

Multiple-Input Multiple-Output Sampling: Necessary Density Conditions

Raman Venkataramani, *Member, IEEE*, and Yoram Bresler, *Fellow, IEEE*

Abstract—We consider the problem of multiple-input multiple-output (MIMO) sampling of multiband signals. In this problem, a set of input signals is passed through a MIMO channel modeled as a known linear time-invariant system. The inputs are modeled as multiband signals whose spectral supports are sets of finite measure and the channel outputs are sampled on nonuniform sampling sets. The aim is to reconstruct the inputs from the output samples. This sampling scheme is quite general and it encompasses various others including Papoulis' generalized sampling and nonuniform sampling as special cases. We introduce notions of joint upper and lower densities for collections of sampling sets and then derive necessary conditions on these densities for stable sampling and consistent reconstruction of the channel inputs from the sampled outputs. These results generalize classical density results for stable sampling and interpolation due to Landau.

Index Terms—Frames, interpolation, multiband signals, multiple-input multiple-output (MIMO) systems, necessary density conditions, reconstruction, stable sampling.

I. INTRODUCTION

MULTICHANNEL deconvolution or multichannel separation of a convolutive mixture is an important problem arising in several applications and has recently attracted substantial interest. In essence, the problem deals with a multiple-input multiple-output (MIMO) channel with observable outputs and the aim is to invert or equalize the channel to recover the channel inputs. In general, the channel inputs have overlapping spectra. Some example applications where MIMO channels arise are multiuser or multiaccess wireless communications and space-time coding with antenna arrays, or telephone digital subscriber loops [1]–[4], multisensor biomedical signals [5], [6], multitrack magnetic recording [7], multiple speaker (or other acoustic source) separation with microphone arrays [8], [9], geophysical data processing [10], and multichannel image restoration [11], [12].

In practice, digital processing is used to perform the channel inversion, whereas the channel inputs and outputs are continuous-time signals. Consequently, the channel outputs are sampled prior to processing. Thus, our aim is to reconstruct the channel inputs from the sampled outputs. This channel inver-

Manuscript received December 17, 2001; revised April 9, 2004. This work was supported in part by a grant from DARPA under Contract F49620-98-1-0498 administered by AFSOR, and by National Science Foundation under Infrastructure Grant CDA-24396. This work was performed while R. Venkataramani was with the University of Illinois at Urbana-Champaign.

R. Venkataramani is with Seagate Technology, Pittsburgh, PA 15222 USA (e-mail: ramanv@ieee.org).

Y. Bresler is with the Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: ybresler@uiuc.edu).

Communicated by G. Battail, Associate Editor at Large.

Digital Object Identifier 10.1109/TIT.2004.831755

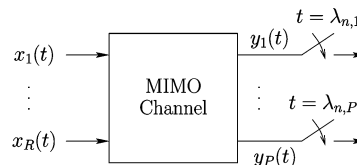


Fig. 1. MIMO sampling.

sion problem can be restated as one in sampling theory, which we call *MIMO sampling*. We study this problem entirely from the perspective of sampling theory, although the problem could, equally well, be viewed as one of channel equalization.

While much of the recent work on MIMO equalization has been on the so-called blind problem, we consider the simpler nonblind problem and assume that the channel characteristics are either known or that they can be estimated accurately using known test input signals. We seek necessary conditions on the sampling sets and the channel for reconstruction of the inputs. These necessary conditions must also apply to the harder problem of blind MIMO equalization with sampling.

Our problem is formulated as follows. Let $x_r(t)$, $r = 1, \dots, R$, be a collection of complex-valued signals whose spectral supports are sets $\mathcal{F}_r \subseteq \mathbb{R}$ of finite measure. We refer to such signals as *multiband* signals because, in practice, \mathcal{F} is a finite union of intervals. These R signals are input to a MIMO channel consisting of linear time-invariant filters (see Fig. 1) to produce P outputs $y_p(t)$, $p = 1, \dots, P$, which may be expressed as

$$y_p(t) = \sum_{r=1}^R g_{pr}(t) * x_r(t), \quad p = 1, \dots, P$$

where $*$ denotes convolution and $\{g_{pr}\}$ are square-integrable impulse responses. Each output $y_p(t)$ is subsequently sampled on a discrete set $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$ and these samples are then used to reconstruct the inputs. This sampling scheme is very general and subsumes various other sampling schemes as special cases. For instance, Papoulis' generalized sampling [13] is essentially a single-input multiple-output (SIMO) sampling scheme, i.e., MIMO sampling with $R = 1$. An extension of Papoulis' sampling expansion to vector-valued inputs [14] is also a special case with all inputs having identical low-pass spectra, i.e., $\mathcal{F}_r = [-B, B]$.

Let $x(t)$, $t \in \mathbb{R}^d$, denote a complex-valued and square-integrable continuous function whose Fourier transform $X(f)$ is supported on a measurable set $\mathcal{F} \subset \mathbb{R}^d$ where d is a positive integer. Landau [15], [16] proved the following fundamental result for sampling and interpolation of multiband signals. Suppose that a function $x(t)$ is sampled on a discrete set $\Lambda = \{\lambda_n :$

$n \in \mathbb{Z} \subseteq \mathbb{R}^d$. Then, for stable reconstruction¹ of $x(t)$ from its samples $x(\lambda_n)$, it is necessary that the sampling density² of Λ be no less than the measure of \mathcal{F} , i.e., Λ must be sufficiently dense in order to stably reconstruct the input. The precise definitions of stable sampling and sampling density are presented in Section II.

A dual problem is that of interpolation: given a discrete set Λ and a square-summable sequence $\{c_n\} \in \ell^2$, we ask if there exists a function $x(t)$ with spectral support \mathcal{F} that interpolates through the values $\{c_n\}$ at the sampling point $\{\lambda_n\}$, i.e., $x(\lambda_n) = c_n$. A necessary condition for interpolation is that the density of Λ be no more than the measure of \mathcal{F} . Roughly speaking, Λ must be sufficiently sparse in order to assign arbitrary values to the samples of $x(t)$ on the set Λ .

We refer to the above problems as classical sampling and interpolation. Gröchenig and Razafinjatovo [17] provided a simpler proof of Landau's classical result for the case that \mathcal{F} has zero boundary measure. Their technique also allowed them to prove necessary density conditions for some derivative sampling schemes. In this paper, we extend the idea of [17] to derive necessary density results for MIMO sampling. We consider only single variate functions in our analysis ($d = 1$), and the results easily extend to multivariate functions. More specifically, we address the following questions. a) What are necessary conditions on $\{\Lambda_p\}$ for stable reconstruction of the MIMO inputs $x_r \in \mathcal{B}(\mathcal{F}_r)$ from the output samples $\{y_p(\lambda_{np})\}$? b) What are necessary conditions on $\{\Lambda_p\}$ such that for any square-summable sequence $\{c_{np}\}$ there exists a set of inputs to the channel whose p th output interpolates through the values $\{c_{np}\}$ at the sampling points $\{\lambda_{np}\}$? We refer to this as *consistent reconstruction*. Problem b) is analogous to the classical interpolation problem. As in the case of classical sampling, our goal is to find necessary conditions on the sampling sets for stable sampling and consistent reconstruction.

In [18], we derived sufficient conditions for reconstruction from MIMO samples assuming that the outputs are sampled uniformly. The related filter design issues were studied in [19]. Sampling theorems for special cases of MIMO sampling are also considered in [13], [14]. These results are sufficient density conditions for uniform or periodic sampling, and are not shown to be necessary for arbitrary, nonuniform sampling of the channel outputs. An interesting SIMO sampling scheme applicable to general signal spaces including wavelet and spline spaces can be found in [20]. However, we restrict our attention to multiband signal spaces alone.

This paper is organized as follows. In Section II, we introduce some notation and review some mathematical background. In Section III, we present our main results: necessary conditions on $\{\Lambda_p\}$ for stable MIMO sampling and consistent MIMO reconstruction. For stable reconstruction, we prove that the sum of densities of Λ_p is lower-bounded by the sum of the measures of \mathcal{F}_r . Similarly, for the consistency problem, the sum of densities of Λ_p is upper-bounded by the sum of the measures of \mathcal{F}_r . Apart from these natural generalizations of the classical results, we also derive necessary conditions on the joint density

for each subcollection of sampling sets, as well as conditions on the channel transfer function. These bounds provide an outer bound on the region of achievable densities. We provide examples to illustrate the results. Finally, we provide proofs of these results in Section IV.

II. PRELIMINARIES

A. Definitions and Notation

Let ℓ^2 denote the set of square-summable sequences and $L^2(\mathbb{R})$ and $C(\mathbb{R})$, the spaces of square-integrable and continuous functions, respectively, on \mathbb{R} . Then

$$\mathcal{B}(\mathcal{F}) = \{x \in L^2(\mathbb{R}) \cap C(\mathbb{R}) : X(f) = 0, \forall f \notin \mathcal{F}\} \quad (1)$$

is the space of continuous $L^2(\mathbb{R})$ signals bandlimited to a measurable set $\mathcal{F} \subseteq \mathbb{R}$, where $X(f)$ is the Fourier transform of a signal $x(t)$

$$X(f) = \int_{\mathbb{R}} x(t)e^{-j2\pi ft} dt.$$

Let $\chi(\cdot)$ denote the indicator function. Thus, $\chi(f \in \mathcal{F})$ takes the value 1 on the set \mathcal{F} , and 0 elsewhere. Let $\phi_{\mathcal{F}}(t)$ denote the inverse Fourier transform of $\chi(f \in \mathcal{F})$

$$\phi_{\mathcal{F}}(t) = \int_{\mathcal{F}} e^{j2\pi ft} df.$$

Let Θ_{τ} denote the time-shift operator: $\Theta_{\tau}f(t) = f(t - \tau)$. Let \emptyset denote the empty set and \mathcal{S}^c , the complement of a set \mathcal{S} in the appropriate universal set. Let $\mu(\cdot)$ denote the Lebesgue measure of real sets.

The space of complex-valued matrices of size $M \times N$ is denoted by $\mathbb{C}^{M \times N}$. Let $\mathbf{e}_r \in \mathbb{C}^{R \times 1}$ denote the r th standard basis vector, i.e., \mathbf{e}_r has a 1 at the r th position, and zeros elsewhere. For a given matrix \mathbf{A} , let \mathbf{A}^H denote its conjugate-transpose $\mathbf{A}_{\mathcal{R}, \mathcal{C}}$, its submatrix corresponding to rows indexed by the set \mathcal{R} and columns by the set \mathcal{C} . Also, let $\mathbf{A}_{\bullet, \mathcal{C}}$ denote the submatrix formed by keeping all rows of \mathbf{A} , but only columns indexed by \mathcal{C} , and $\mathbf{A}_{\mathcal{R}, \bullet}$, the submatrix formed by retaining rows indexed by \mathcal{R} , and all columns. We use a similar notation for vectors. Thus, $\mathbf{X}_{\mathcal{R}}$ is the subvector of \mathbf{X} corresponding to rows indexed by \mathcal{R} . When dealing with singleton index sets: $\mathcal{R} = \{r\}$ or $\mathcal{C} = \{c\}$, we omit the curly braces for readability. Therefore, $\mathbf{A}_{r, \bullet}$ and $\mathbf{A}_{\bullet, c}$ are the r th row and the c th column of \mathbf{A} , respectively. Let $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote the largest and smallest eigenvalues of \mathbf{A} . Recall that the singular values of a matrix \mathbf{A} are the nonzero eigenvalues of $\sqrt{\mathbf{A}^H \mathbf{A}}$. Let $\sigma_{\max}(\mathbf{A})$ denote the largest singular value of matrix \mathbf{A} , and $\sigma_{\min}(\mathbf{A})$, the smallest *nonzero* singular value of \mathbf{A} if $\mathbf{A} \neq \mathbf{0}$. If $\mathbf{A} = \mathbf{0}$, we take $\sigma_{\min}(\mathbf{A}) = \infty$.

B. Sampling Density

A discrete set $\Lambda = \{\lambda_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}$ is called *uniformly discrete* with separation δ if

$$|\lambda_m - \lambda_n| \geq 2\delta, \quad \forall m \neq n.$$

¹The property that any errors in the sample values cause a controlled amount of error in the reconstruction.

²Interpreted as the average number of samples per unit time.

Let the maximum and minimum number of sampling points of Λ found in any interval of length 2γ be denoted by

$$\begin{aligned}\nu_\gamma^+(\Lambda) &= \sup_{\tau \in \mathbb{R}} \#(\Lambda \cap B_\gamma(\tau)) \\ \nu_\gamma^-(\Lambda) &= \inf_{\tau \in \mathbb{R}} \#(\Lambda \cap B_\gamma(\tau))\end{aligned}\quad (2)$$

respectively, where $\#(\mathcal{S})$ denotes the cardinality of a set \mathcal{S} , and

$$B_\gamma(\tau) = \{\sigma \in \mathbb{R} : |\sigma - \tau| \leq \gamma\}$$

is a closed interval of length 2γ centered at τ . For a discrete set Λ , the upper and lower densities are defined as

$$\begin{aligned}D^+(\Lambda) &= \limsup_{\gamma \rightarrow \infty} \frac{\nu_\gamma^+(\Lambda)}{2\gamma} \\ D^-(\Lambda) &= \liminf_{\gamma \rightarrow \infty} \frac{\nu_\gamma^-(\Lambda)}{2\gamma}\end{aligned}\quad (3)$$

respectively. Although traditionally written as “lim inf” and “lim sup,” the limits in (3) can be replaced by simple limits [21]. Several other notions of density for nonuniform sampling can be found in [21].

If the lower and upper densities coincide, this density is called the *uniform density*³ and is denoted by $D(\Lambda)$. Any large interval of length l contains approximately $lD(\Lambda)$ points of Λ . If Λ is uniformly discrete, then $D^+(\Lambda)$ is finite but the converse statement is not true. However, $D^+(\Lambda) < \infty$ implies that Λ can be expressed as a union of finitely many uniformly discrete sets [22].

When dealing with a collection of sampling sets, as in the MIMO setting, it is useful to define *joint* densities for the collection. In [22], we introduced the following generalizations of sampling density.

Definition 1: Given a finite collection of discrete sets Λ_p , $p = 1, \dots, P$, their joint upper and lower densities are defined as

$$D^+(\Lambda_1, \dots, \Lambda_P) = \limsup_{\gamma \rightarrow \infty} \frac{\nu_\gamma^+(\Lambda_1, \dots, \Lambda_P)}{2\gamma}\quad (4)$$

$$D^-(\Lambda_1, \dots, \Lambda_P) = \liminf_{\gamma \rightarrow \infty} \frac{\nu_\gamma^-(\Lambda_1, \dots, \Lambda_P)}{2\gamma}\quad (5)$$

respectively, where

$$\nu_\gamma^+(\Lambda_1, \dots, \Lambda_P) = \sup_{\tau \in \mathbb{R}} \sum_{p=1}^P \#(\Lambda_p \cap B_\gamma(\tau))$$

$$\nu_\gamma^-(\Lambda_1, \dots, \Lambda_P) = \inf_{\tau \in \mathbb{R}} \sum_{p=1}^P \#(\Lambda_p \cap B_\gamma(\tau))$$

are the maximum and minimum number of sampling points of the collection $\{\Lambda_p : p = 1, \dots, P\}$ found in any interval of length 2γ .

Once again, the limits in (4) and (5) can be replaced with simple limits. In fact, this can be inferred from the following stronger result.

Proposition 1: For all $\gamma > 0$

$$\nu_\gamma^+(\Lambda_1, \dots, \Lambda_P) \geq 2\gamma D^+(\Lambda_1, \dots, \Lambda_P)\quad (6)$$

$$\nu_\gamma^-(\Lambda_1, \dots, \Lambda_P) \leq 2\gamma D^-(\Lambda_1, \dots, \Lambda_P).\quad (7)$$

³The sampling points in Λ need not be uniformly spaced.

Proof: Suppose that $d^+ = D^+(\Lambda_1, \dots, \Lambda_P) < \infty$. Let $\epsilon > 0$ and $\gamma > 0$ be arbitrary. Then (4) guarantees the existence of $\tau \in \mathbb{R}$ and $\gamma' \geq \gamma/\epsilon$ such that

$$\sum_{p=1}^P \#(\Lambda_p \cap B_{\gamma'}(\tau)) \geq 2\gamma'(d^+ - \epsilon).$$

Let $n \in \mathbb{Z}$ be such that $\gamma'/\gamma \leq n < \gamma'/\gamma + 1$. Suppose we divide the interval $\mathcal{I} = B_{\gamma'}(\tau)$ into n equal intervals of width $2\gamma'/n$, then it is clear that for at least one interval, say \mathcal{I}_0 , we have

$$\sum_{p=1}^P \#(\Lambda_p \cap \mathcal{I}_0) \geq 2\gamma'(d^+ - \epsilon)/n.$$

Therefore, by choice of γ' and n

$$\begin{aligned}\sum_{p=1}^P \#(\Lambda_p \cap \mathcal{I}_0) &\geq \frac{2\gamma'(d^+ - \epsilon)}{1 + \gamma'/\gamma} = \frac{2\gamma(d^+ - \epsilon)}{1 + \gamma/\gamma'} \\ &\geq \frac{2\gamma(d^+ - \epsilon)}{1 + \epsilon}.\end{aligned}$$

Since $m(\mathcal{I}_0) = 2\gamma'/n \leq 2\gamma$, it follows that

$$\nu_\gamma^+(\Lambda_1, \dots, \Lambda_P) \geq \frac{2\gamma(d^+ - \epsilon)}{1 + \epsilon}$$

for all $\gamma > 0$, and we obtain (6) because $\epsilon > 0$ is arbitrary. If $d^+ = \infty$, then for every $\epsilon > 0$, we can find $\gamma' \geq \gamma/\epsilon$ such that

$$\sum_{p=1}^P \#(\Lambda_p \cap B_{\gamma'}(\tau)) \geq 2\gamma'/\epsilon.$$

Proceeding as before, we find that there exists an interval \mathcal{I}_0 such that $m(\mathcal{I}_0) \leq 2\gamma$ and

$$\sum_{p=1}^P \#(\Lambda_p \cap \mathcal{I}_0) \geq \frac{2\gamma}{\epsilon(1 + \epsilon)}.$$

Since $\epsilon > 0$ is arbitrary, we obtain $\nu_\gamma^+(\Lambda_1, \dots, \Lambda_P) = \infty$. The proof of (7) is very similar. \square

Proposition 1 implies that

$$\liminf_{\gamma \rightarrow \infty} \frac{\nu_\gamma^+(\Lambda_1, \dots, \Lambda_P)}{2\gamma} \geq D^+(\Lambda_1, \dots, \Lambda_P)$$

$$\liminf_{\gamma \rightarrow \infty} \frac{\nu_\gamma^-(\Lambda_1, \dots, \Lambda_P)}{2\gamma} \leq D^-(\Lambda_1, \dots, \Lambda_P).$$

Thus, the limits in (4) and (5) can be replaced by simple limits.

C. Stable Sampling

In classical sampling, samples of a signal $x \in \mathcal{B}(\mathcal{F})$ on a discrete set Λ are used to reconstruct x . The input space is the Hilbert space $\mathcal{H} = \mathcal{B}(\mathcal{F})$ with the following inner product:

$$\langle x, y \rangle = \int_{\mathbb{R}} x(t)\overline{y(t)}dt, \quad \forall x, y \in \mathcal{B}(\mathcal{F}).$$

The norm on \mathcal{H} is defined as $\|x\| = \sqrt{\langle x, x \rangle}$. Thus, the sampling operation can be expressed as an inner product

$$x(\lambda) = \langle x, \Theta_\lambda \phi_{\mathcal{F}} \rangle.$$

The set $\Lambda = \{\lambda_n : n \in \mathbb{Z}\}$ is called a set of stable sampling for $\mathcal{B}(\mathcal{F})$ if there exist $A, B > 0$ such that

$$A\|\mathbf{x}\|^2 \leq \sum_{n \in \mathbb{Z}} |x(\lambda_n)|^2 \leq B\|\mathbf{x}\|^2, \quad \forall \mathbf{x} \in \mathcal{B}(\mathcal{F}). \quad (8)$$

This condition implies that $\{\Theta_{\lambda_n} \phi_{\mathcal{F}} : n \in \mathbb{Z}\}$ is a frame for $\mathcal{B}(\mathcal{F})$ with frame bounds A and B and the condition number $K = \sqrt{B/A} \geq 1$. The theory of frames thus provides a convenient tool to study sampling [21]. We present more details on frames in Section IV.

We extend the above definition of stable sampling to the MIMO problem as follows. Recall that the channel inputs and outputs are related to each other as

$$\mathbf{y}(t) = \mathbf{g}(t) * \mathbf{x}(t) = \int_{\mathbb{R}} \mathbf{g}(t - \tau) \mathbf{x}(\tau) d\tau$$

where \mathbf{x} is the input vector whose components are multiband signals $x_r \in \mathcal{B}(\mathcal{F}_r)$, and \mathbf{y} is the channel output in vector form, and \mathbf{g} is a matrix whose entries are $g_{pr}(t)$. Thus, the space of inputs is

$$\mathcal{H} = \mathcal{B}(\mathcal{F}_1) \times \cdots \times \mathcal{B}(\mathcal{F}_R).$$

Suppose that

$$\mathcal{C}_f = \{r : f \in \mathcal{F}_r\}. \quad (9)$$

Then it is clear that $\mathbf{X}_{\mathcal{C}_f}(f)$ contains all the nonzero elements of $\mathbf{X}(f)$. Hence, the channel output can be expressed in the frequency domain as

$$\mathbf{Y}(f) = \mathbf{G}(f)\mathbf{X}(f) = \mathbf{G}_{\bullet, \mathcal{C}_f}(f)\mathbf{X}_{\mathcal{C}_f}(f) \quad (10)$$

where $\mathbf{G}(f)$ (the Fourier transform of $\mathbf{g}(t)$) is called the channel transfer function matrix. We use the shorthand notation $\mathbf{y} = \mathbf{G}\mathbf{x}$ to denote the input–output relation of the channel, where G is the operator denoting the channel.

In the rest of this paper let

$$\mathcal{R} = \{1, \dots, R\} \quad \text{and} \quad \mathcal{P} = \{1, \dots, P\}$$

denote index sets for the channel inputs and outputs, respectively.

Definition 2: A collection of discrete sampling sets $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$, $p \in \mathcal{P}$ is said to be *stable* with respect to (G, \mathcal{H}) if there exist $A, B > 0$ such that

$$A\|\mathbf{x}\|^2 \leq \sum_{p=1}^P \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 \leq B\|\mathbf{x}\|^2 \quad (11)$$

for every $\mathbf{x} \in \mathcal{H}$, where $\mathbf{y} = \mathbf{G}\mathbf{x}$. We sometimes refer to this as a collection of stable MIMO sampling.

As we will see later, the implication of this definition is that the inputs to the channel can be reconstructed from the samples of the outputs on these discrete sets in a stable way, i.e., any error in the sampled values produces a controlled error in the reconstruction.

D. Interpolation and Consistency

In the context of classical multiband sampling, a sampling set Λ is called a set of interpolation if for every $\{c_n\} \in l^2$ there exists $x \in \mathcal{B}(\mathcal{F})$ such that $x(\lambda_n) = c_n$, i.e., the sampling operator

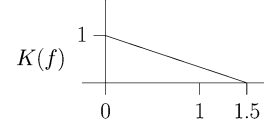


Fig. 2. $K(f)$.

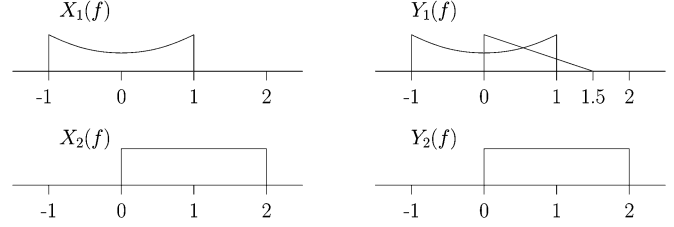


Fig. 3. Typical spectra of the channel inputs and outputs.

corresponding to Λ from $\mathcal{B}(\mathcal{F})$ to l^2 is onto if Λ is a set of interpolation. There are several practical implications to Λ being a set of interpolation. First, it implies that any square-summable data sequence $\{c_n\}$ can be interpolated to a signal $x \in \mathcal{B}(\mathcal{F})$ whose samples on Λ agree with the data sequence. Second, it implies that the samples of $x \in \mathcal{B}(\mathcal{F})$ on Λ are nonredundant because each sample is completely independent of all the others. If Λ is not a set of interpolation, then the samples of $x \in \mathcal{B}(\mathcal{F})$ live in a strict subspace of l^2 , and are linearly dependent with some samples completely determined by the others. We extend this notion to MIMO sampling as follows.

Definition 3: A collection of discrete sets $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$, $p \in \mathcal{P}$ is said to be *consistent* with respect to (G, \mathcal{H}) if for every $\{c_{np}\} \in l^2$, there exists a solution $\mathbf{x} \in \mathcal{H}$ to the problem $y_p(\lambda_{np}) = c_{np}$, where $\mathbf{y} = \mathbf{G}\mathbf{x}$. We sometimes refer to this as a collection of consistent MIMO reconstruction.

In the preceding definition, we seek an input signal that, when passed through the MIMO channel and sampled, produces the desired observations. We refer to this property as *consistent reconstruction* (as opposed to *interpolation*) because the observations are samples of the channel output (as opposed to the inputs).

III. NECESSARY DENSITY CONDITIONS

In this section we present our main results namely, necessary conditions for stable sampling and consistent reconstruction. We begin with an example for illustration.

Example 1: Consider a MIMO channel with $R = 2$ inputs, $P = 2$ outputs, and the following transfer function matrix:

$$\mathbf{G}(f) = \begin{bmatrix} 1 & K(f) \\ 0 & 1 \end{bmatrix} \quad (12)$$

where $K(f) = (1 - 2f/3)\chi_{f \in [0, 1.5]}$ is shown in Fig. 2. Let $\mathcal{F}_1 = [-1, 1)$ and $\mathcal{F}_2 = [0, 2)$ be the input spectral supports. Thus,

$$y_1(t) = x_1(t) + k(t) * x_2(t) \quad (13)$$

$$y_2(t) = x_2(t). \quad (14)$$

The input and output spectra for a typical set of channel inputs are illustrated in Fig. 3. We interpret $Y_1(f)$ as the sum of the two pieces shown in the figure.

Let $y_1(t)$ and $y_2(t)$ be sampled on sets Λ_1 and Λ_2 , respectively. Then, what are necessary conditions on Λ_1 and Λ_2 for stable reconstruction of the inputs?

We shall first find a necessary condition on Λ_1 assuming that $y_2(t)$ is known for all $t \in \mathbb{R}$ instead of $t \in \Lambda_2$. This is allowed because the resulting condition will also be necessary for the original problem. Thus, $x_2(t) = y_2(t)$ by (14). Using (13) we have

$$y_1(\lambda_{n1}) = x_1(\lambda_{n1}) + [k(t) * x_2(t)]_{t=\lambda_{n1}}. \quad (15)$$

Since $x_2(t)$ is known for all t , the second term on the right-hand side can be computed and subtracted off from the left-hand side (known samples) to yield the samples of $x_1(t)$ on Λ_1

$$x_1(\lambda_{n1}) = y_1(\lambda_{n1}) - [k(t) * x_2(t)]_{t=\lambda_{n1}}. \quad (16)$$

Given $x_2(t)$, the data contained in the sequences (15) and (16) are equivalent in the sense that one can be used to find the other. Therefore, we must be able to reconstruct $x_1(t)$ from $\{x_1(\lambda_{n1})\}$ and $x_2(t)$. Since $x_1(t)$ and $x_2(t)$ are independent of each other, $x_1(t)$ must be reconstructed from the sequence (16) alone. Using Landau's result, we obtain the following necessary condition on Λ_1 :

$$D^-(\Lambda_1) \geq \mu(\mathcal{F}_1) = 2.$$

Next, we determine a necessary condition on Λ_2 . Note that if we can reconstruct $x_1(t)$ and $x_2(t)$, then outputs $y_1(t)$ and $y_2(t)$ are also trivially reconstructible. This time, we assume that $y_1(t)$ is known for all $t \in \mathbb{R}$ and any resulting necessary condition on Λ_2 for stable reconstruction of $y_2(t)$ from $y_1(t)$ and $\{y_2(\lambda_{n2}) : n \in \mathbb{Z}\}$ will also be necessary for the original problem.

Let $\mathcal{Z} = [1, 1.5]$. From Fig. 3 it is clear that for $f \in \mathcal{Z}$

$$Y_1(f) = K(f)X_2(f) = K(f)Y_2(f).$$

Therefore, the part of the spectrum of $Y_2(f)$, $f \in \mathcal{Z}$ can be recovered (although not stably because $K(f)$ approaches zero at $f = 1.5$) from $Y_1(f)$. However, no part of $Y_2(f)$, $f \notin \mathcal{Z}$ can be recovered from $Y_1(f)$. This observation motivates us to decompose the signal $y_2(t)$ as

$$y_2(t) = y_a(t) + y_b(t)$$

where $Y_a(f) = Y_2(f)\chi(f \in \mathcal{Z})$ and $Y_b(f) = Y_2(f)\chi(f \notin \mathcal{Z})$. Then, $y_a(t)$ can be recovered from $y_1(t)$. Having determined $y_a(t)$, we compute the samples of $y_b(t)$ on Λ_2 as follows:

$$y_b(\lambda_{n2}) = y_2(\lambda_{n2}) - y_a(\lambda_{n2}).$$

Using a similar argument as before, we obtain the following condition for stable reconstruction of $y_b(t)$ (and hence $y_2(t)$):

$$D^-(\Lambda_2) \geq \mu(\mathcal{F}_2 \cap \mathcal{Z}^c) = 1.5.$$

Finally, we expect the total sampling density of the outputs to be larger than the total spectral measure of the inputs, i.e.,

$$D^-(\Lambda_1, \Lambda_2) \geq \mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) = 4.$$

In the above example, we used a series of steps to deduce necessary conditions on the sampling densities. However, our arguments were not very rigorous but specific to the given transfer function $G(f)$. The necessary density conditions for the general

problem can be stated in a very simple form. We postpone their proofs to Section IV since they are technically involved.

A. Density Conditions for Stable Sampling

Let the essential supremum and infimum of a real function $A(f)$ be defined as

$$\begin{aligned} \text{ess sup } A(f) &= \inf\{a : A(f) \leq a \text{ a.e.}\} \\ \text{ess inf } A(f) &= \sup\{a : A(f) \geq a \text{ a.e.}\} \end{aligned}$$

where a.e. stands for *almost everywhere*.

Theorem 1: Suppose that \mathcal{F}_r , $r \in \mathcal{R}$, are real sets of finite measure, $\mathcal{H} = \mathcal{B}(\mathcal{F}_1) \times \cdots \times \mathcal{B}(\mathcal{F}_R)$, and Λ_p , $p \in \mathcal{P}$, are discrete sets with $D^+(\Lambda_p) < \infty$ that constitute a collection of stable sampling with respect to (G, \mathcal{H}) . Then, for every $\Pi \subseteq \mathcal{P}$

$$D^-(\{\Lambda_p : p \in \Pi\}) \geq \sum_{r=1}^R \mu(\mathcal{F}_r) - \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)) df \quad (17)$$

where $\mathcal{C}_f = \{r : f \in \mathcal{F}_r\}$ and Π^c is the complement of Π in \mathcal{P} . Furthermore, if

$$\text{ess inf}_{f \in \mathcal{F}} \sigma_{\min}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)) = 0, \quad \mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r \quad (18)$$

for some $\Pi \neq \mathcal{P}$, then (17) is a strict inequality.

Theorem 1, which is proved in Section IV, provides lower bounds on the joint densities of all subcollections of $\{\Lambda_p\}$. In particular, letting $\Pi = \mathcal{P}$ in (17), we obtain

$$D^-(\Lambda_1, \dots, \Lambda_P) \geq \sum_{r=1}^R \mu(\mathcal{F}_r). \quad (19)$$

In other words, the combined sampling density on all the output channels must be no less than the combined bandwidth of all the input signals, which represents the total number of degrees of freedom per unit time contained in the inputs.

Intuitively, we can explain these bounds as follows. Suppose that the outputs $y_p(t)$, $p \in \Pi^c$ are completely known for all $t \in \mathbb{R}$, which is the case that demands the weakest conditions from the $\{\Lambda_p : p \in \Pi\}$ for stable sampling. Then, $\mathbf{Y}_{\Pi^c}(f) = \mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)\mathbf{X}_{\mathcal{C}_f}(f)$ is known for all f . Therefore, $\text{rank}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f))$ is the number of independent components of $\mathbf{X}(f)$ at frequency f that can be determined from knowledge of $\mathbf{Y}_{\Pi^c}(f)$ alone. Consequently

$$\int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)) df$$

is the number of degrees of freedom per unit time in the inputs that can be resolved by knowing the outputs $y_p(t)$, $p \in \Pi^c$ completely (for all t). Therefore, the difference in right-hand side of (17) is the number of unresolved degrees of freedom per unit time in the inputs. The left-hand side of (17) is the joint lower density of $\{\Lambda_p : p \in \Pi\}$, i.e., the smallest local sampling density (number of samples per unit time in a local sense) contained in these sampling sets. Thus, (17) merely states that we require more samples than the unresolved degrees of freedom in the inputs (locally per unit time) for each choice of Π .

Note that this bound depends only on the submatrix of $\mathbf{G}(f)$ whose rows are indexed by the complement of Π and columns by \mathcal{C}_f because $X_r(f)$ vanishes outside \mathcal{F}_r .

Next, if some singular value of $\mathbf{G}_{\Pi^c, c_f}(f)$ takes arbitrarily small nonzero values, then we cannot stably invert $\mathbf{Y}_{\Pi^c}(f) = \mathbf{G}_{\Pi^c, c_f}(f)\mathbf{X}_{c_f}(f)$ to stably recover the independent components of $\mathbf{X}(f)$ and the density of Λ_p must be strictly larger than the right-hand side of (17).

Theorem 1 leads to the following simple necessary conditions on the *admissibility* of subsets of the continuous-time channel outputs for stable recovery of the inputs. Let $\mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r$.

Definition 4: A set of outputs $y_p, p \in \Pi, \Pi \subseteq \mathcal{P}$ is said to be an *admissible* set of outputs for \mathcal{H} if

$$\text{ess inf}_{f \in \mathcal{F}} \lambda_{\min}(\mathbf{G}_{\Pi, c_f}^H(f)\mathbf{G}_{\Pi, c_f}(f)) > 0. \quad (20)$$

It is easily verified that (20), which states that the singular values of $\mathbf{G}_{\Pi, c_f}(f)$ are uniformly bounded away from zero, is a necessary condition for stable recovery of $\mathbf{x} \in \mathcal{H}$ from the continuous-time outputs $\{y_p(t) : t \in \mathbb{R}, p \in \Pi\}$.

Corollary 1: Under the hypotheses of Theorem 1, Π^c is an admissible output set for \mathcal{H} for every $\Pi \subset \mathcal{P}, \Pi \neq \mathcal{P}$, for which $D^-(\{\Lambda_p : p \in \Pi\}) = 0$. In particular

$$\text{ess inf}_{f \in \mathcal{F}} \lambda_{\min}(\mathbf{G}_{\bullet, c_f}^H(f)\mathbf{G}_{\bullet, c_f}(f)) > 0. \quad (21)$$

Proof: From Theorem 1, (42), and (43) we have

$$\begin{aligned} D^-(\{\Lambda_p : p \in \Pi\}) &\geq \sum_{r=1}^R \mu(\mathcal{F}_r) - \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi^c, c_f}(f)) df \\ &\geq \sum_{r=1}^R \mu(\mathcal{F}_r) - \int_{\mathbb{R}} |c_f| df \\ &\geq 0. \end{aligned}$$

If $D^-(\{\Lambda_p : p \in \Pi\}) = 0$, then all inequalities above must be equalities. Thus, (17) holds with an equality, implying that

$$\text{ess inf}_{f \in \mathcal{F}} \sigma_{\min}(\mathbf{G}_{\Pi^c, c_f}(f)) > 0$$

by Theorem 1. We also have $\text{rank}(\mathbf{G}_{\Pi^c, c_f}(f)) = |c_f|$ a.e., implying that

$$\lambda_{\min}(\mathbf{G}_{\Pi^c, c_f}^H(f)\mathbf{G}_{\Pi^c, c_f}(f)) = [\sigma_{\min}(\mathbf{G}_{\Pi^c, c_f}(f))]^2.$$

Now, the admissibility of Π^c follows by combining the last two observations. Applying this result to $\Pi = \emptyset$, we obtain (21). \square

Equation (21), which states that the entire set \mathcal{P} of outputs must be admissible for stable MIMO sampling is not surprising: even if all $y_p(t)$ are known for $t \in \mathbb{R}$, we cannot stably recover the channel inputs unless (21) holds. In fact, an even simpler necessary condition emerges from (21)

$$P \geq |c_f| \text{ a.e.}$$

i.e., the number of outputs must be no less than the number of overlapping input spectra at any frequency.

Next, suppose that $D^-(\{\Lambda_p : p \in \Pi\}) = 0$. Then, the output samples on the sampling sets $\{\Lambda_p : p \in \Pi\}$ are too sparse to contain any signal information. Therefore, we must rely entirely on the output samples taken on $\{\Lambda_p : p \in \Pi^c\}$ to achieve stable reconstruction, and an argument as before provides intuitive justification for the admissibility of Π^c .

The following result provides another necessary condition for stable sampling.

Theorem 2: Under the hypotheses of Theorem 1, we require

$$\text{ess sup}_{f \in \mathcal{F}} \sigma_{\max}(\mathbf{G}_{\Pi^+, c_f}(f)) < \infty \quad (22)$$

where $\Pi^+ = \{p \in \mathcal{P} : D^+(\Lambda_p) > 0\}$ and $\mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r$.

Theorem 2 is proved in Section IV. Whenever $D^+(\Lambda_p) = 0$ for some $p \in \mathcal{P}$, the samples of y_p on Λ_p are too sparse to provide any useful information. Thus, Π^+ can be viewed as the set of outputs whose samples are dense enough to provide information about the inputs. In view of this interpretation, we see that (22) is an implication of the upper stability bound in (11).

Clearly, $D^-(\Lambda_p) = D^+(\Lambda_p) = 0$ for all $p \notin \Pi^+$. Thus, $D^-(\{\Lambda_p : p \in (\Pi^+)^c\}) = 0$ and by Corollary 1, Π^+ must be an admissible set and (20) holds for $\Pi = \Pi^+$. In this case, the condition (20) applied to Π^+ and condition (22) are necessary bounds on the smallest and largest singular values of the same matrix. So, if (22) does not hold, then stable reconstruction from the sampled versions outputs will be impossible even if Π^+ is admissible. An example of this in the single-input single-output case is an integrator $\mathbf{G}(f) = 1/(j2\pi f)$.

We now present a simple example to illustrate the necessary conditions for stable MIMO sampling.

Example 2: Consider the MIMO channel and inputs as described in Example 1. We seek conditions on the sampling sets Λ_1 and Λ_2 for stable MIMO sampling with respect to G . We have $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 = [-1, 2]$ and

$$\mathbf{C}_f = \{r : f \in \mathcal{F}_r\} = \begin{cases} \{1\}, & \text{if } f \in [-1, 0) \\ \{1, 2\}, & \text{if } f \in [0, 1) \\ \{2\}, & \text{if } f \in [1, 2) \\ \emptyset, & \text{otherwise.} \end{cases}$$

It is easy to check that (22) is satisfied regardless of Π^+ . Also,

$$\sigma_{\min}(\mathbf{G}_{\bullet, c_f}(f)) = \frac{[2 + K^2(f)] - \sqrt{[2 + K^2(f)]^2 - 4}}{2}.$$

This quantity is positive and uniformly bounded away from zero because $K(f)$ is a bounded function. Hence, (21) is satisfied. Applying Theorem 1, we obtain the following density conditions:

$$\begin{aligned} D^-(\Lambda_1, \Lambda_2) &\geq \mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) = 4 \\ D^-(\Lambda_1) &\geq \mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) - \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{2, c_f}(f)) df \\ &= 4 - \int_{[-1, 0]} 0 df - \int_{[0, 1]} 1 df - \int_{[1, 2]} 1 df = 2 \\ D^-(\Lambda_2) &\geq \mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) - \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{1, c_f}(f)) df \\ &= 4 - \int_{[-1, 0]} 1 df \\ &\quad - \int_{[0, 1]} 1 df - \int_{[1, 2]} \chi(K(f) \neq 0) df \\ &= 1.5. \end{aligned}$$

Now, a simple calculation reveals that

$$\sigma_{\min}(\mathbf{G}_{1, c_f}(f)) = \begin{cases} 1, & \text{if } f \in [-1, 0) \\ \sqrt{1 + K^2(f)}, & \text{if } f \in [0, 1) \\ |K(f)|, & \text{if } f \in [1, 2). \end{cases}$$

Clearly, this quantity takes arbitrarily small values in the vicinity of $f = 1.5$, where $K(f)$ vanishes. Hence, the bound on $D^-(\Lambda_2)$ is a strict inequality. Another calculation yields⁴

$$\sigma_{\min}(\mathbf{G}_{2, \mathcal{C}_f}(f)) = \begin{cases} \infty, & \text{if } f \in [-1, 0) \\ 1, & \text{if } f \in [0, 1) \\ 1, & \text{if } f \in [1, 2). \end{cases}$$

Hence, the bound on $D^-(\Lambda_1)$ is not a strict inequality. In summary, we obtain the following necessary conditions on the joint densities:

$$D^-(\Lambda_1, \Lambda_2) \geq 4, \quad D^-(\Lambda_1) \geq 2, \quad \text{and} \quad D^-(\Lambda_2) > 1.5.$$

These results agree with our predictions in Example 1.

Finally, note that we can have undersampling at each output and yet be able to reconstruct all the inputs jointly from the available information. For instance, we do not need $D^-(\Lambda_1) \geq 2.5$, even though $y_1(t)$ has a bandwidth of 2.5. To see this, we construct a sampling scheme for which the densities $(d_1, d_2) = (2, 2)$ are *achievable*, where $d_p = D(\Lambda_p)$, i.e., Λ_p has a uniform density of d_p . Let Λ_1 and Λ_2 be uniform sampling lattices

$$\Lambda_1 = \Lambda_2 = \{n/2 : n \in \mathbb{Z}\}.$$

Clearly, $y_2(t) = x_2(t)$ can be reconstructed stably. Now, the samples of $x_1(t)$ on Λ_1 can be computed as follows:

$$x_1(\lambda_{n1}) = y_1(\lambda_{n1}) - [k(t) * x_2(t)]_{t=\lambda_{n1}}$$

because $x_2(t)$ is known for all t . Thus, $x_1(t)$ can also be reconstructed stably. However, it is not immediately clear whether all densities satisfying the necessary conditions are achievable or how to achieve them.

B. Density Conditions for Consistent Reconstruction

We now present the necessary condition for consistent MIMO reconstruction, which is dual to the problem of stable sampling.

Theorem 3: Suppose that \mathcal{F}_r , $r \in \mathcal{R}$ are real sets of finite measure, $\mathcal{H} = \mathcal{B}(\mathcal{F}_1) \times \dots \times \mathcal{B}(\mathcal{F}_R)$, and Λ_p , $p \in \mathcal{P}$ are discrete sets that constitute a collection of consistent reconstruction with respect to (G, \mathcal{H}) . Then, for every $\Pi \subseteq \mathcal{P}$

$$D^+(\{\Lambda_p : p \in \Pi\}) \leq \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) df \quad (23)$$

where $\mathcal{C}_f = \{r : f \in \mathcal{F}_r\}$. Furthermore, if

$$\text{ess inf}_{f \in \mathcal{F}} \sigma_{\min}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) = 0, \quad \mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r \quad (24)$$

for some $\Pi \neq \emptyset$, then (23) is a strict inequality.

Theorem 3 is proved in Section IV. Suppose $\mathbf{G}(f)$ satisfies (21), then

$$\text{rank}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) \leq |\mathcal{C}_f|.$$

Under this condition, $\Pi = \mathcal{P}$ in (23) implies that

$$D^+(\Lambda_1, \dots, \Lambda_P) \leq \sum_{r=1}^R \mu(\mathcal{F}_r) \quad (25)$$

i.e., the joint density of Λ_p does not exceed the combined bandwidth of the input signals. Note that (21) need not hold for consistent reconstruction.

⁴Recall that we take $\sigma_{\min}(\mathbf{A}) = \infty$ if $\mathbf{A} = 0$.

Theorem 3 provides conditions on the joint upper densities of all subcollections of $\{\Lambda_p\}$. The density bounds can be interpreted as follows. The right-hand side of (23) is the joint upper density of the sampling sets $\{\Lambda_p : p \in \Pi\}$, i.e., the largest local sampling density (number of samples per unit time in a local sense) in these sampling sets. We have already seen that the quantity

$$\int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) df$$

is the number of degrees of freedom per unit time contained in the outputs $\{y_p(t) : p \in \Pi\}$. Thus, (23) states that in order to interpolate the outputs through its given sample values, the local density of samples must be less than the number of degrees of freedom per unit time in these outputs. The following corollary is dual to Corollary 1.

Corollary 2: Under the hypotheses of Theorem 3, Π is an admissible output set for \mathcal{H} , i.e.,

$$\text{ess inf}_{f \in \mathcal{F}} \lambda_{\min}(\mathbf{G}_{\Pi, \mathcal{C}_f}^H(f) \mathbf{G}_{\Pi, \mathcal{C}_f}(f)) > 0 \quad (26)$$

for every $\Pi \subseteq \mathcal{P}$, $\Pi \neq \emptyset$, such that

$$D^+(\{\Lambda_p : p \in \Pi\}) = \sum_{r=1}^R \mu(\mathcal{F}_r). \quad (27)$$

Proof: From Theorem 3 and (42), we have

$$\begin{aligned} D^+(\{\Lambda_p : p \in \Pi\}) &\leq \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) df \\ &\leq \int_{\mathbb{R}} |\mathcal{C}_f| = \sum_{r=1}^R \mu(\mathcal{F}_r). \end{aligned}$$

If (27) holds, then both the above inequalities are, in fact, equalities. Then, (23) is satisfied with an equality, implying that (24) fails to hold. Also, $\text{rank}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) = |\mathcal{C}_f|$, so that

$$\lambda_{\min}(\mathbf{G}_{\Pi, \mathcal{C}_f}^H(f) \mathbf{G}_{\Pi, \mathcal{C}_f}(f)) = [\sigma_{\min}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f))]^2.$$

Now (26) follows by combining the last two observations. \square

Corollary 2 can be interpreted as follows. Suppose that the smallest singular value of $\mathbf{G}_{\Pi, \mathcal{C}_f}(f)$ takes arbitrarily small values, then there are unit energy inputs that produce outputs $y_p(t)$, $p \in \Pi$, of arbitrarily small energies. Roughly speaking, we can find l^2 sequences $\{c_{np}\}$ for which the consistency problem does not have a finite-energy solution $\mathbf{x} \in \mathcal{H}$ if we are operating at the critical sampling density, i.e., with an equality in (23).

Example 3: Let the MIMO channel and the input spectral supports be as defined in Example 1. What are the necessary conditions on the sampling sets Λ_1 and Λ_2 for consistent reconstruction? Fortunately, we have already performed the necessary calculations in Example 2. Applying Theorem 3, we obtain the following bounds on the joint densities:

$$\begin{aligned} D^+(\Lambda_1, \Lambda_2) &\leq \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\bullet, \mathcal{C}_f}(f)) df = 4 \\ D^+(\Lambda_1) &< \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{1, \mathcal{C}_f}(f)) df = 2.5 \\ D^+(\Lambda_2) &\leq \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{2, \mathcal{C}_f}(f)) df = 2. \end{aligned}$$

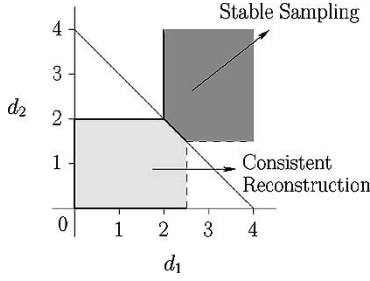


Fig. 4. Density regions for stable sampling and consistent reconstruction.

These inequalities can be explained intuitively as follows. If consistent reconstructions of $x_1(t)$ and $x_2(t)$ are possible, then we must also have consistent reconstruction (or interpolation) of $y_1(t)$ and $y_2(t)$ from their respective samples. Looking at Fig. 3, we see that $Y_2(f) = X_2(f)$ is bandlimited to $[0, 2)$. Therefore, we require $D^+(\Lambda_2) \leq 2$ for consistent reconstruction of $y_2(t)$. Next, $Y_1(f)$ is bandlimited to $[-1, 2.5)$, thereby requiring $D^+(\Lambda_1) \leq 2.5$. However, $D^+(\Lambda_1) = 2.5$ is not allowed for stability reasons because $K(f)$ is arbitrarily small in the vicinity of $f = 1.5$. Finally, the combined bandwidth of the inputs is 4, so that $D^+(\Lambda_1, \Lambda_2) \leq 4$ is needed for consistency.

We now show that the densities $D(\Lambda_1) = D(\Lambda_2) = 2$ for sampling sets $\{\Lambda_p\}$ of uniform density. Let Λ_1 and Λ_2 be defined as in Example 2. Let $\{c_{n1}\}$ and $\{c_{n2}\}$ be l^2 sequences. Then, the problem $y_2(\lambda_{n2}) = c_{n2}$ clearly has a solution $y_2 \in \mathcal{B}([0, 2])$. Now, the sequence $d_n = [k(t) * x_2(t)]_{t=\lambda_{n1}}$ is square-summable because $K(f)$ is a bounded function, implying that

$$x_1(\lambda_{n1}) = y_1(\lambda_{n1}) - [k(t) * x_2(t)]_{t=\lambda_{n1}} = c_{n1} - d_n$$

also has a solution $x_1 \in \mathcal{B}([-1, 1])$. This proves that $(d_1, d_2) = (2, 2)$ is achievable.

If Λ_1 and Λ_2 have uniform densities of d_1 and d_2 , respectively, the resulting outer bounds on the density regions for stable sampling (Example 2) and consistent reconstruction (Example 3) can be viewed as regions in \mathbb{R}^2 . These regions are illustrated in Fig. 4.

IV. PROOFS OF MAIN RESULTS

This section is devoted to the proofs of the results of Section III. We begin with a review of background material on frames and bases (cf. [23], [24]). Let \mathcal{H} be a separable Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$.

Definition 5: A sequence $\{\psi_n\} \subseteq \mathcal{H}$ is called a *frame* if there exist $A, B > 0$ such that

$$A\|x\|^2 \leq \sum_n |\langle \psi_n, x \rangle|^2 \leq B\|x\|^2 \quad (28)$$

for all $x \in \mathcal{H}$. The constants A and B are called the lower and upper *frame bounds*. If $A = B$, the frame is a tight frame.

The frame operator $S : \mathcal{H} \rightarrow \mathcal{H}$ defined by

$$Sx = \sum_n \langle x, \psi_n \rangle \psi_n, \quad \forall x \in \mathcal{H}$$

is a bounded linear operator satisfying $AI \leq S \leq BI$, where I is the identity operator. Let $\tilde{\psi}_n = S^{-1}\psi_n$. Then, $\{\tilde{\psi}_n\}$ is also a frame (the *dual frame*) for \mathcal{H} with frame bounds B^{-1} and A^{-1} . Any $x \in \mathcal{H}$ can be expanded as

$$x = \sum_n \langle x, \tilde{\psi}_n \rangle \psi_n = \sum_n \langle x, \psi_n \rangle \tilde{\psi}_n. \quad (29)$$

Also, for any sequence $\{c_n\} \in l^2$, we have

$$\left\| \sum_n c_n \psi_n \right\|^2 \leq B \sum_n |c_n|^2. \quad (30)$$

Definition 6: A sequence $\{\psi_n\} \subseteq \mathcal{H}$ is called a *Riesz basis* if it is fully equivalent to an orthonormal basis for \mathcal{H} , i.e., if there exists a bounded invertible operator T and an orthonormal basis $\{e_n\}$ such that $\psi_n = Te_n$.

A Riesz basis is a frame, and hence (29) and (30) hold. In fact, if $\{\psi_n\}$ is a Riesz basis, we can replace (30) by the stronger condition

$$A \sum_n |c_n|^2 \leq \left\| \sum_n c_n \psi_n \right\|^2 \leq B \sum_n |c_n|^2. \quad (31)$$

Conversely, if $\{\psi_n\}$ is a complete sequence in \mathcal{H} , it is a Riesz basis for \mathcal{H} whenever (31) holds for finite sequences [23]. The dual frame $\{\tilde{\psi}_n\}$ of a Riesz basis $\{\psi_n\}$ is also a Riesz basis for \mathcal{H} and is called the *biorthogonal basis* of $\{\psi_n\}$.

Definition 7: A sequence $\{\psi_n\} \subseteq \mathcal{H}$ is called a *Riesz–Fischer sequence* if the *moment problem*:

$$\langle x, \psi_n \rangle = c_n \quad (32)$$

has a solution $x \in \mathcal{H}$ for all $\{c_n\} \in l^2$. If $\{\psi_n\}$ is a Riesz–Fischer sequence, there exists a solution x to (32) such that

$$\|x\|^2 \leq \frac{1}{a} \|c\|^2$$

for some $a > 0$ called the bound of the Riesz–Fischer sequence.

A necessary and sufficient condition for $\{\psi_n\}$ to be a Riesz–Fischer sequence with bound a is that

$$\left\| \sum_n c_n \psi_n \right\|^2 \geq a \sum_n |c_n|^2 \quad (33)$$

for every finite sequence $\{c_n\}$. Finally, note that the moment problem in (32) has a unique solution if $\{\psi_n\}$ is a *complete* Riesz–Fischer sequence in \mathcal{H} . Every Riesz basis is a Riesz–Fischer sequence, but the converse is not true. However, if a Riesz–Fischer sequence is also a frame, then it is a Riesz basis. We use the notions of frames and Riesz–Fischer sequences in our analysis in this section.

The space of input signals in the MIMO channel is the Hilbert space

$$\mathcal{H} = \mathcal{B}(\mathcal{F}_1) \times \cdots \times \mathcal{B}(\mathcal{F}_R) \quad (34)$$

equipped with the inner product

$$\langle \mathbf{x}, \mathbf{z} \rangle = \int_{\mathbb{R}} \mathbf{z}^H(t) \mathbf{x}(t) dt = \sum_{r=1}^R \int_{\mathbb{R}} \overline{z_r(t)} x_r(t) dt \quad (35)$$

where $\mathbf{x}, \mathbf{z} \in \mathcal{H}$. The norm on \mathcal{H} is $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. It is clear that we can write $y_p(\lambda_{np}) = \langle \mathbf{x}, \Theta_{\lambda_{np}} \boldsymbol{\psi}_p \rangle$ where $\boldsymbol{\psi}_p \in \mathcal{H}$ is given by

$$\boldsymbol{\Psi}_p(f) = \sum_{r=1}^R \overline{G_{pr}(f)} \chi(f \in \mathcal{F}_r) \mathbf{e}_r \quad (36)$$

and \mathbf{e}_r is the r th standard basis vector.

Thus, (11) is equivalent to the condition that

$$\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : n \in \mathbb{Z}, p \in \mathcal{P}\}$$

is a frame for \mathcal{H} . This implies that the channel inputs can be reconstructed from the output samples using the dual frame as the set of interpolating functions. Furthermore, any errors in the sampled signal or its samples produce a controlled amount of error in the reconstructed signals. The *condition number* for the MIMO sampling scheme is $K = \sqrt{B/A} \geq 1$.

Next, Definition 3 implies that $\{\Lambda_p : p \in \mathcal{P}\}$ is a collection of consistent reconstructions with respect to (G, \mathcal{H}) if

$$\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : n \in \mathbb{Z}, p \in \mathcal{P}\}$$

is a Riesz–Fischer sequence in \mathcal{H} . Note that for any finite sequence $\{c_{np}\}$

$$\begin{aligned} \left\| \sum_{n,p} c_{np} \Theta_{\lambda_{np}} \boldsymbol{\psi}_p \right\|^2 &= \max_{\mathbf{x} \in B_{\mathcal{H}}} \left| \sum_{n,p} c_{np} \langle \mathbf{x}, \Theta_{\lambda_{np}} \boldsymbol{\psi}_p \rangle \right|^2 \\ &= \max_{\mathbf{x} \in B_{\mathcal{H}}} \left| \sum_{n,p} c_{np} y_p(\lambda_{np}) \right|^2 \end{aligned}$$

where $B_{\mathcal{H}} = \{\mathbf{x} \in \mathcal{H} : \|\mathbf{x}\| \leq 1\}$ is the unit ball in \mathcal{H} . From (33) and the above observation we conclude that

$$\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : n \in \mathbb{Z}, p \in \mathcal{P}\}$$

is a Riesz–Fischer sequence in \mathcal{H} if and only if

$$\max_{\mathbf{x} \in B_{\mathcal{H}}} \left| \sum_{n,p} c_{np} y_p(\lambda_{np}) \right|^2 \geq a \sum_{n,p} |c_{np}|^2 \quad (37)$$

for every finite sequence $\{c_{np}\}$. This characterization (37) is more convenient to use than Definition 3.

Finally, we point out that if a collection of discrete sets $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$, $p \in \mathcal{P}$ is a collection of both stable sampling and consistent reconstruction, then $\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : n \in \mathbb{Z}, p \in \mathcal{P}\}$ is a Riesz basis for \mathcal{H} .

A. Preliminary Results

Our aim is to prove necessary density conditions for stable sampling and consistent reconstruction stated in Section III. These results are analogous to Landau's density result for nonuniform sampling of multiband signals [15], [16]. Gröchenig and Razafinjatoivo [17] provided a simpler proof of Landau's result for signals with spectral sets having zero boundary measure. We extend their idea to prove our density results for the MIMO problem. However, we drop the restriction on the boundary measures.

We begin with a few relevant definitions. Let

$$\begin{aligned} \mathcal{H}_\beta &= (\mathcal{B}([-\beta, \beta]))^R, \quad \beta > 0 \\ \mathcal{H}_\infty &= (L^2(\mathbb{R}))^R \end{aligned}$$

be Hilbert spaces with inner product defined as in (35). The elements of \mathcal{H}_β are vectors whose components are bandlimited to

the frequencies $[-\beta, \beta]$. Let $P_{\mathcal{S}} : \mathcal{H}_\infty \rightarrow \mathcal{S}$ denote the orthogonal projection operator onto a closed subspace $\mathcal{S} \subseteq \mathcal{H}_\infty$.

Definition 8: A subspace $\mathcal{S} \subseteq \mathcal{H}_\infty$ is called *shift-invariant* if $\Theta_\sigma \mathbf{x} \in \mathcal{S}$ for all $\sigma \in \mathbb{R}$, $\mathbf{x} \in \mathcal{S}$.

Evidently \mathcal{H}_∞ and \mathcal{H}_β are shift-invariant spaces. We write $\mathbf{x} \perp \mathcal{S}$ if $\langle \mathbf{x}, \mathbf{z} \rangle = 0$ for all $\mathbf{z} \in \mathcal{S}$. The following properties of a closed shift-invariant subspace $\mathcal{S} \subseteq \mathcal{H}_\infty$ can be verified easily.

Proposition 2: Suppose that $\mathbf{x} \in \mathcal{H}_\beta$, $\sigma \in \mathbb{R}$, and $\mathcal{S} \subseteq \mathcal{H}_\infty$ is a closed shift-invariant subspace. Then (a) $\mathbf{x} \perp \mathcal{S} \implies \Theta_\sigma \mathbf{x} \perp \mathcal{S}$, and (b) $P_{\mathcal{S}} \Theta_\sigma \mathbf{x} = \Theta_\sigma P_{\mathcal{S}} \mathbf{x}$, i.e., translation commutes with orthogonal projection onto \mathcal{S} .

Proof: (a) Suppose that $\mathbf{x} \perp \mathcal{S}$. Then $\langle \Theta_\sigma \mathbf{x}, \mathbf{z} \rangle = \langle \mathbf{x}, \Theta_{-\sigma} \mathbf{z} \rangle = 0$ for all $\mathbf{z} \in \mathcal{S}$, $\sigma \in \mathbb{R}$ because $\Theta_{-\sigma} \mathbf{z} \in \mathcal{S}$. Hence, $\Theta_\sigma \mathbf{x} \perp \mathcal{S}$. To prove (b), note that $P_{\mathcal{S}} \Theta_\sigma \mathbf{x} - \Theta_\sigma P_{\mathcal{S}} \mathbf{x} \in \mathcal{S}$. For an arbitrary $\mathbf{z} \in \mathcal{S}$, we have

$$\begin{aligned} \langle P_{\mathcal{S}} \Theta_\sigma \mathbf{x} - \Theta_\sigma P_{\mathcal{S}} \mathbf{x}, \mathbf{z} \rangle &= \langle P_{\mathcal{S}} \Theta_\sigma \mathbf{x}, \mathbf{z} \rangle - \langle \Theta_\sigma P_{\mathcal{S}} \mathbf{x}, \mathbf{z} \rangle \\ &= \langle \Theta_\sigma \mathbf{x}, P_{\mathcal{S}} \mathbf{z} \rangle - \langle \mathbf{x}, P_{\mathcal{S}} \Theta_{-\sigma} \mathbf{z} \rangle \\ &= \langle \Theta_\sigma \mathbf{x}, \mathbf{z} \rangle - \langle \mathbf{x}, \Theta_{-\sigma} \mathbf{z} \rangle = 0. \end{aligned}$$

Thus, $P_{\mathcal{S}} \Theta_\sigma \mathbf{x} = \Theta_\sigma P_{\mathcal{S}} \mathbf{x}$. \square

The following theorem is a stronger version of the main result in [17]. This result allows us to compute necessary density conditions for the stable MIMO sampling and consistent reconstruction.

Theorem 4 (Comparison Theorem): Let \mathcal{H}_S and \mathcal{H}_L be closed subspaces of \mathcal{H}_∞ , and let $\Sigma_1, \dots, \Sigma_Q$ and $\Lambda_1, \dots, \Lambda_P$ be discrete subsets of \mathbb{R} such that all $D^+(\Lambda_p) < \infty$. Suppose that $\mathbf{s}_1, \dots, \mathbf{s}_Q$ and $\mathbf{l}_1, \dots, \mathbf{l}_P$ are elements of \mathcal{H}_∞ such that

$$\{\Theta_\sigma \mathbf{s}_q : \sigma \in \Sigma_q, q = 1, \dots, Q\} \subseteq \mathcal{H}_S$$

is a Riesz–Fischer sequence in \mathcal{H}_S , and that

$$\{\Theta_\lambda \mathbf{l}_p : \lambda \in \Lambda_p, p = 1, \dots, P\} \subseteq \mathcal{H}_L$$

is a frame for \mathcal{H}_L . Then

$$D^\pm(\Lambda_1, \dots, \Lambda_P) \geq D^\pm(\Sigma_1, \dots, \Sigma_Q) - \sum_{q=1}^Q \alpha_q D^+(\Sigma_q) \quad (38)$$

where

$$\alpha_q = \frac{1}{\sqrt{a}} \sup_{\sigma \in \Sigma_q} \|\Theta_\sigma \mathbf{s}_q - P_{\mathcal{H}_L} \Theta_\sigma \mathbf{s}_q\|.$$

Furthermore, $D^+(\Sigma_q) < \infty$ is automatically implied whenever all $\alpha_q < 1$.

Theorem 4 is proved in [22]. Note that \mathcal{H}_L and \mathcal{H}_S are arbitrary subspaces in \mathcal{H}_∞ . However, the comparison theorem is most useful when the spaces are nearly the same. In this case, the coefficients α_q would be small, thereby yielding the following density bound:

$$D^\pm(\Lambda_1, \dots, \Lambda_P) \geq D^\pm(\Sigma_1, \dots, \Sigma_Q) - \epsilon$$

where $\epsilon > 0$ is a small quantity representing the summation in (38) involving the terms α_q . By using an appropriate limiting argument, we would need to show that ϵ can be made arbitrarily small. The import of this statement is roughly that a frame, being an overcomplete sequence in a Hilbert space \mathcal{H} , is denser (contains more vectors) than a Riesz–Fischer sequence \mathcal{H} .

Note that Theorem 4 is very general, involving arbitrary signal spaces, and can potentially be used for proving necessary density conditions for sampling problems in other spaces such as wavelet or spline spaces. We use this theorem in the next section, where we derive necessary density conditions for the MIMO sampling problem. Finally, we state a useful result proved in [22].

Lemma 1: Let $h \in \mathcal{B}([\nu_1, \nu_2])$. Then

$$h^\#(t) \stackrel{\text{def}}{=} \sup_{|\tau-t| \leq 1} |h(\tau)|$$

satisfies $h^\# \in L^2(\mathbb{R})$ and $\|h^\#\|^2 \leq C\|h\|^2$ for some $C = C(\nu_2 - \nu_1) > 0$ that depends only on the difference $(\nu_2 - \nu_1)$. Moreover, if $\Lambda = \{\lambda_m : m \in \mathbb{Z}\} \subseteq \mathbb{R}$ is a discrete set with $D^+(\Lambda) < \infty$, then

$$\sum_{|\lambda_n - \sigma| \geq \Gamma} |h(\lambda_n)|^2 \leq C' \int_{|t-\sigma| \geq \Gamma-1} |h^\#(t)|^2 dt \quad (39)$$

for all $\sigma \in \mathbb{R}$, $\Gamma \geq 0$, and some $C' = C'(\Lambda) > 0$. In particular

$$\sum_{n \in \mathbb{Z}} |h(\lambda_n)|^2 \leq C' C \|h\|^2.$$

Lemma 1 says that the samples of a unit-energy bandlimited signal on a sampling set of finite upper density cannot have arbitrarily large energy.

B. Proof of Theorem 1

Let $\Pi \subseteq \mathcal{P}$ be a fixed subset. Consider two cases: first suppose that $\bar{\Pi} \neq \mathcal{P}$ and (18) holds. Then, we define

$$K = \max_{p \in \Pi^c} C'(\Lambda_p) C(1) \quad (40)$$

where C' and C are quantities defined in Lemma 1. Let $\epsilon_0 > 0$ be such that $K\epsilon_0^2 \leq A/2$, where A is the lower stability bound in (11). Since (18) is satisfied, there exists a set \mathcal{D}_0 such that $\mu(\mathcal{D}_0) > 0$ and

$$\sigma_{\min}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)) \leq \epsilon_0, \quad \forall f \in \mathcal{D}_0. \quad (41)$$

Without loss of generality, assume that $\mathcal{D}_0 \subseteq [\nu, \nu+1]$ for some $\nu \in \mathbb{R}$. In the second case (either $\bar{\Pi} = \mathcal{P}$ or (18) does not hold), we take $\mathcal{D}_0 = \emptyset$. Thus, (41) is satisfied in both cases. Let the cardinality of the set \mathcal{C}_f be denoted by $|\mathcal{C}_f|$, i.e.,

$$|\mathcal{C}_f| = \sum_{r=1}^R \chi(f \in \mathcal{F}_r). \quad (42)$$

Let the dimension of the null space of $\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)$ be denoted by

$$\rho(f) = |\mathcal{C}_f| - \text{rank}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)) \quad (43)$$

and let the columns of $\mathbf{U}'(f) \in \mathbb{C}^{|\mathcal{C}_f| \times \rho(f)}$ form an orthonormal basis for this space. For $f \in \mathcal{D}_0$, let $\mathbf{U}''(f) \in \mathbb{C}^{|\mathcal{C}_f| \times 1}$ be a unit-norm right singular vector of $\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)$ corresponding to its smallest nonzero singular value. We can always choose $\mathbf{U}'(f)$ and $\mathbf{U}''(f)$ to be measurable functions. Clearly, $\mathbf{U}''(f)$ is orthogonal to the columns of $\mathbf{U}'(f)$ for $f \in \mathcal{D}_0$. Therefore,

$$\mathbf{U}(f) \stackrel{\text{def}}{=} \begin{cases} [\mathbf{U}'(f) \ \mathbf{U}''(f)], & \text{if } f \in \mathcal{D}_0 \\ \mathbf{U}'(f), & \text{otherwise} \end{cases}$$

has orthonormal columns for all f . Let \mathcal{G}_r be the set where $\mathbf{U}(f)$ contains r columns, i.e.,

$$\mathcal{G}_r = \{f : \rho(f) + \chi(f \in \mathcal{D}_0) = r\}, \quad r \in \mathcal{R}. \quad (44)$$

The sets $\{\mathcal{G}_r\}$ are clearly disjoint sets of finite measure. Therefore, for any $\delta > 0$, there exists a finite collection of disjoint intervals $\{\mathcal{I}_{rk} : r \in \mathcal{R}, k = 1, \dots, K_r\}$ such that the sets

$$\mathcal{G}'_r \stackrel{\text{def}}{=} \bigcup_{k=1}^{K_r} \mathcal{I}_{rk}, \quad r \in \mathcal{R} \quad (45)$$

approximate \mathcal{G}_r in the sense that $\mu(\mathcal{G}'_r \cap \mathcal{G}_r^c) \leq \delta/R^2$ and $\mu(\mathcal{G}_r^c \cap \mathcal{G}'_r) \leq \delta/R^2$. It follows that

$$\sum_{r=1}^R r \mu(\mathcal{G}'_r \cap \mathcal{G}_r^c) \leq \sum_{r=1}^R \frac{r\delta}{R^2} \leq \delta. \quad (46)$$

It is also clear that $|\mu(\mathcal{G}'_r) - \mu(\mathcal{G}_r)| \leq \delta/R^2$. Consequently

$$\left| \sum_{r=1}^R r \mu(\mathcal{G}_r) - \sum_{r=1}^R r \mu(\mathcal{G}'_r) \right| \leq \delta. \quad (47)$$

Now, define $\mathbf{W}^r(f) \in \mathbb{C}^{R \times r}$ on \mathcal{G}'_r for each r as follows:

$$\mathbf{W}^r(f) = \begin{cases} \mathbf{I}_{\bullet, \mathcal{C}_f} \mathbf{U}(f), & \text{if } f \in \mathcal{G}'_r \cap \mathcal{G}_r \\ (\mathbf{e}_1, \dots, \mathbf{e}_r), & \text{if } f \in \mathcal{G}'_r \cap \mathcal{G}_r^c \end{cases} \quad (48)$$

where \mathbf{I} is the $R \times R$ identity matrix. Note that the columns of $\mathbf{W}^r(f)$ form an orthonormal set of vectors for each $f \in \mathcal{G}'_r$. For each $r \in \mathcal{R}$, let k and m be indexes such that $1 \leq k \leq K_r$ and $1 \leq m \leq r$. For convenience, let $q(r, k, m)$ denote an invertible mapping from the triplet (r, k, m) to a single index q

$$q : \{(r, k, m) : r \in \mathcal{R}, k = 1, \dots, K_r, m = 1, \dots, r\} \rightarrow \mathcal{Q}$$

where $\mathcal{Q} = \{1, \dots, Q\}$ and

$$Q = \sum_{r=1}^R r K_r.$$

In the rest of the proof, assume that q, r, k , and m are related to each other by $q = q(r, k, m)$. We shall now define several quantities with the intention of finally using Theorem 4 to derive the necessary density conditions. Let $\{\mathbf{s}_q\} \subseteq \mathcal{H}_\infty$ be defined in terms of their Fourier transforms as follows:

$$\mathbf{s}_q(f) = \begin{cases} \mathbf{W}_{\bullet, m}^r(f) / \sqrt{\mu(\mathcal{I}_{rk})}, & \text{if } f \in \mathcal{I}_{rk} \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (49)$$

where $\mathbf{W}_{\bullet, m}^r(f)$ is the m th column of $\mathbf{W}^r(f)$. The sampling set

$$\Sigma_q \stackrel{\text{def}}{=} \left\{ \frac{n}{\mu(\mathcal{I}_{rk})} : n \in \mathbb{Z} \right\}, \quad \forall m \in \{1, \dots, r\} \quad (50)$$

has uniform density of $\mu(\mathcal{I}_{rk})$. Since the intervals \mathcal{I}_{rk} are disjoint and $\{\mathbf{W}_{\bullet, m}^r(f) : m = 1, \dots, r\}$ is a set of orthonormal vectors for each r and f , it follows that

$$\{\Theta_{\sigma_{nq}} \mathbf{s}_q : q \in \mathcal{Q}, n \in \mathbb{Z}\}$$

is an orthonormal sequence. Let \mathcal{H}_S be the closure of the span of this sequence

$$\mathcal{H}_S = \overline{\text{span}\{\Theta_{\sigma_{nq}} \mathbf{s}_q : q \in \mathcal{Q}, n \in \mathbb{Z}\}} \subseteq \mathcal{H}_\infty. \quad (51)$$

Then $\{\Theta_{\sigma_{nq}} \mathbf{s}_q : q \in \mathcal{Q}, n \in \mathbb{Z}\}$ is an orthonormal Riesz basis for \mathcal{H}_S with lower frame bound $a = 1$. In particular, it is a Riesz–Fischer sequence with bound $a = 1$.

Now define

$$\mathcal{H}_L = \{\mathbf{x} \in \mathcal{H} : \mathbf{X}_{C_f}(f) = \mathbf{U}(f)\mathbf{U}^H(f)\mathbf{X}_{C_f}(f) \text{ a.e.}\}. \quad (52)$$

Clearly, \mathcal{H}_L is a shift-invariant subspace. To see that \mathcal{H}_L is closed, consider the following argument. Let $\{\mathbf{x}^j\} \in \mathcal{H}_L$ be a sequence converging to $\mathbf{x}^\infty \in \mathcal{H}_\infty$. Then

$$\mathbf{X}_{C_f}^i(f) = \mathbf{U}(f)\mathbf{U}^H(f)\mathbf{X}_{C_f}^i(f) \text{ a.e.}$$

and $\mathbf{X}^i(f)$ converges to $\mathbf{X}^\infty(f)$ in the L^2 sense. Hence, there exists a subsequence $\{i_j\}$ such that $\mathbf{X}^{i_j}(f) \rightarrow \mathbf{X}^\infty(f)$ a.e. as $j \rightarrow \infty$. Therefore,

$$\begin{aligned} \mathbf{U}(f)\mathbf{U}^H(f)\mathbf{X}_{C_f}^\infty(f) &= \lim_{j \rightarrow \infty} \mathbf{U}(f)\mathbf{U}^H(f)\mathbf{X}_{C_f}^{i_j}(f) \\ &= \mathbf{X}_{C_f}^\infty(f) \text{ a.e.} \end{aligned}$$

or, equivalently, $\mathbf{x}_\infty \in \mathcal{H}_L$, proving that \mathcal{H}_L is closed.

Suppose that $\mathbf{x} \in \mathcal{H}_L$. Then, using (10) and (52) we see that

$$\begin{aligned} \mathbf{Y}_{\Pi^c}(f) &= \mathbf{G}_{\Pi^c, C_f}(f)\mathbf{X}_{C_f}(f) \\ &= \mathbf{G}_{\Pi^c, C_f}(f)\mathbf{U}(f)\mathbf{U}^H(f)\mathbf{X}_{C_f}(f). \end{aligned}$$

Next, using the definitions of $\mathbf{U}(f)$ and $\mathbf{U}''(f)$, we obtain

$$\mathbf{Y}_{\Pi^c}(f) = \mathbf{0}, \quad f \notin \mathcal{D}_0 \quad (53)$$

$$\begin{aligned} \|\mathbf{Y}_{\Pi^c}(f)\| &= \|\mathbf{G}_{\Pi^c, C_f}(f)\mathbf{U}''(f)\mathbf{U}''^H(f)\mathbf{X}_{C_f}(f)\| \\ &\leq \epsilon_0 \|\mathbf{X}(f)\|, \quad f \in \mathcal{D}_0. \end{aligned} \quad (54)$$

Equations (53) and (54) imply that $\|\mathbf{Y}_{\Pi^c}(f)\| \leq \epsilon_0 \|\mathbf{X}(f)\|$ for all f . Hence,

$$\int_{\mathbb{R}} \|\mathbf{Y}_{\Pi^c}(f)\|^2 df \leq \epsilon_0^2 \|\mathbf{x}\|^2. \quad (55)$$

Equation (53) states that $Y_p(f)$ is supported on $\mathcal{D}_0 \subseteq [\nu, \nu + 1]$ for each $p \in \Pi^c$. Applying Lemma 1 to $y_p, p \in \Pi^c$ and using (55) we get

$$\sum_{p \in \Pi^c} \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 \leq K \int_{\mathbb{R}} \|\mathbf{Y}_{\Pi^c}(f)\|^2 df \leq K \epsilon_0^2 \|\mathbf{x}\|^2 \quad (56)$$

where K is the constant defined in (40). Combining (56) with the first inequality in the sampling stability condition in (11), we obtain

$$\begin{aligned} \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 &\geq (A - K \epsilon_0^2) \|\mathbf{x}\|^2 \\ &\geq \frac{A}{2} \|\mathbf{x}\|^2, \quad \forall \mathbf{x} \in \mathcal{H}_L \end{aligned} \quad (57)$$

where the second inequality in (57) follows from the choice of ϵ_0 . From (11), we obviously also have

$$\sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 \leq B \|\mathbf{x}\|^2, \quad \forall \mathbf{x} \in \mathcal{H}_L \quad (58)$$

because $\mathcal{H}_L \subseteq \mathcal{H}$. Equations (57) and (58) yield

$$\frac{A}{2} \|\mathbf{x}\|^2 \leq \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 \leq B \|\mathbf{x}\|^2, \quad \forall \mathbf{x} \in \mathcal{H}_L. \quad (59)$$

Let $\mathbf{l}_p = P_{\mathcal{H}_L} \boldsymbol{\psi}_p$, where $\boldsymbol{\psi}_p$ is defined in (36). Since \mathcal{H}_L is shift-invariant, we obtain the following using Proposition 2:

$$\Theta_{\lambda_{np}} \mathbf{l}_p = \Theta_{\lambda_{np}} P_{\mathcal{H}_L} \boldsymbol{\psi}_p = P_{\mathcal{H}_L} \Theta_{\lambda_{np}} \boldsymbol{\psi}_p \in \mathcal{H}_L. \quad (60)$$

Also,

$$\begin{aligned} \langle \mathbf{x}, \Theta_{\lambda_{np}} \mathbf{l}_p \rangle &= \langle \mathbf{x}, P_{\mathcal{H}_L} \Theta_{\lambda_{np}} \boldsymbol{\psi}_p \rangle \\ &= \langle \mathbf{x}, \Theta_{\lambda_{np}} \boldsymbol{\psi}_p \rangle = y_p(\lambda_{np}), \quad \forall \mathbf{x} \in \mathcal{H}_L. \end{aligned} \quad (61)$$

From (59)–(61) we conclude that $\{\Theta_{\lambda_{np}} \mathbf{l}_p : p \in \Pi, n \in \mathbb{Z}\}$ is a frame for \mathcal{H}_L . Having verified all the required hypotheses, we can now apply Theorem 4 to obtain the following inequality relating the densities of $\{\Lambda_p : p \in \Pi\}$ and $\{\Sigma_q : q \in \mathcal{Q}\}$:

$$D^-(\{\Lambda_p : p \in \Pi\}) \geq D^-(\Sigma_1, \dots, \Sigma_Q) - \sum_{q \in \mathcal{Q}} \alpha_q D^+(\Sigma_q) \quad (62)$$

where

$$\alpha_q = \frac{1}{\sqrt{a}} \sup_{\sigma \in \Sigma_q} \|\Theta_\sigma \mathbf{s}_q - P_{\mathcal{H}_L} \Theta_\sigma \mathbf{s}_q\|.$$

Since \mathcal{H}_L is shift-invariant, we use Proposition 2 to obtain

$$\begin{aligned} \alpha_q &= \frac{1}{\sqrt{a}} \sup_{\sigma \in \Sigma_q} \|\Theta_\sigma \mathbf{s}_q - \Theta_\sigma P_{\mathcal{H}_L} \mathbf{s}_q\| \\ &= \frac{1}{\sqrt{a}} \|\mathbf{s}_q - P_{\mathcal{H}_L} \mathbf{s}_q\| = \|\mathbf{s}_q - P_{\mathcal{H}_L} \mathbf{s}_q\| \end{aligned}$$

where the last equality holds because $a = 1$. In order to estimate α_q , we define $\mathbf{v}_q \in \mathcal{H}$ as follows:

$$\mathbf{V}_q(f) = \begin{cases} \mathbf{S}_q(f), & \text{if } f \in \mathcal{I}_{rk} \cap \mathcal{G}_r \\ \mathbf{0}, & \text{otherwise.} \end{cases} \quad (63)$$

For all $f \in \mathcal{I}_{rk} \cap \mathcal{G}_r$, we use (48), (49), and (63) to conclude that

$$\begin{aligned} \mathbf{U}(f)\mathbf{U}^H(f)\mathbf{V}_q(f) &= \frac{\mathbf{U}(f)\mathbf{U}^H(f)\mathbf{U}_{\bullet, m}(f)}{\sqrt{\mu(\mathcal{I}_{rk})}} \\ &= \frac{\mathbf{U}_{\bullet, m}(f)}{\sqrt{\mu(\mathcal{I}_{rk})}} = \mathbf{V}_q(f). \end{aligned}$$

This proves that $\mathbf{v}_q \in \mathcal{H}_L$. Therefore, $\alpha_q \leq \|\mathbf{s}_q - \mathbf{v}_q\|$. Using Parseval's theorem, (49), and (63) we obtain

$$\begin{aligned} \alpha_q &\leq \left(\int \|\mathbf{S}_q(f) - \mathbf{V}_q(f)\|^2 df \right)^{\frac{1}{2}} \\ &= \left(\int_{\mathcal{I}_{rk} \cap \mathcal{G}_r^c} \left\| \frac{\mathbf{W}_{\bullet, m}^r(f)}{\sqrt{\mu(\mathcal{I}_{rk})}} \right\|^2 df \right)^{\frac{1}{2}}. \end{aligned}$$

Since $\mathbf{W}_{\bullet, m}^r(f)$ is a vector of unit norm, we have

$$\alpha_q \leq \sqrt{\frac{\mu(\mathcal{I}_{rk} \cap \mathcal{G}_r^c)}{\mu(\mathcal{I}_{rk})}}. \quad (64)$$

Combining (62) and (64) and using the fact that $D(\Sigma_q) = \mu(\mathcal{I}_{rk})$, we obtain

$$\begin{aligned} D^-(\{\Lambda_p : p \in \Pi\}) &\geq \sum_{r, m, k} \mu(\mathcal{I}_{rk}) - \sum_{r, m, k} \alpha_q \mu(\mathcal{I}_{rk}) \\ &\geq \sum_r r \mu(\mathcal{G}_r') \\ &\quad - \sum_{r, k} r \sqrt{\mu(\mathcal{I}_{rk} \cap \mathcal{G}_r^c) \mu(\mathcal{I}_{rk})}. \end{aligned}$$

Using the Cauchy–Schwarz inequality, we obtain

$$D^-(\{\Lambda_p : p \in \Pi\}) \quad (65)$$

$$\begin{aligned} &\geq \sum_r r\mu(\mathcal{G}'_r) - \left(\sum_{rk} r\mu(\mathcal{I}_{rk} \cap \mathcal{G}'_r) \right)^{\frac{1}{2}} \left(\sum_{rk} r\mu(\mathcal{I}_{rk}) \right)^{\frac{1}{2}} \\ &= \sum_r r\mu(\mathcal{G}'_r) - \left(\sum_r r\mu(\mathcal{G}'_r \cap \mathcal{G}^c_r) \right)^{\frac{1}{2}} \left(\sum_r r\mu(\mathcal{G}'_r) \right)^{\frac{1}{2}}. \end{aligned} \quad (66)$$

Now, (46), (47), and (66) imply that

$$\begin{aligned} D^-(\{\Lambda_p : p \in \Pi\}) &\geq \left[\sum_r r\mu(\mathcal{G}_r) \right] - \delta \\ &\quad - \left[\delta + \sum_r r\mu(\mathcal{G}_r) \right]^{\frac{1}{2}}. \end{aligned} \quad (67)$$

Meanwhile, using (42)–(44), and the definition of the Lebesgue integral, we obtain

$$\begin{aligned} \sum_{r=1}^R r\mu(\mathcal{G}_r) &= \int_{\mathbb{R}} [\rho(f) + \chi(f \in \mathcal{D}_0)] df \\ &= \mu(\mathcal{D}_0) + \int_{\mathbb{R}} [|\mathcal{C}_f| - \text{rank}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f))] df \\ &= \mu(\mathcal{D}_0) + \sum_{r=1}^R \mu(\mathcal{F}_r) - \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)) df. \end{aligned} \quad (68)$$

Putting together (67) and (68), and letting $\delta \rightarrow 0$ yields

$$\begin{aligned} D^-(\{\Lambda_p : p \in \Pi\}) &\geq \mu(\mathcal{D}_0) + \sum_{r=1}^R \mu(\mathcal{F}_r) \\ &\quad - \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi^c, \mathcal{C}_f}(f)) df. \end{aligned}$$

This proves (17). Finally, recall that if (18) holds and $\Pi \neq \mathcal{P}$, then $\mu(\mathcal{D}_0) > 0$. Thus, the inequality in (17) is strict in this case. \square

C. Proof of Theorem 2

Suppose that (22) fails to hold. Then, some entry of $\mathbf{G}_{\Pi^+, \mathcal{C}_f}(f)$ is necessarily unbounded on \mathcal{F} . So, let $p_o \in \Pi^+$ and $r_o \in \mathcal{R}$ be indexes such that for every $\epsilon > 0$ there exists

$$\mathcal{G} \subseteq \mathcal{F}_{r_o} \cap \{f : |G_{p_o, r_o}(f)|^2 \geq 1/\epsilon\}$$

satisfying $0 < \mu(\mathcal{G}) < \infty$. Let $\{\mathcal{I}_k : k = 1, \dots, K\}$ be a finite collection of disjoint intervals such that

$$\mathcal{G}' = \bigcup_{k=1}^K \mathcal{I}_k$$

satisfies $\mu(\mathcal{G}' \cap \mathcal{G}^c) \leq \delta$ and $\mu(\mathcal{G}'^c \cap \mathcal{G}) \leq \delta$, where $\delta = \epsilon\mu(\mathcal{G})/(1 + \epsilon)$. Therefore, $\mu(\mathcal{G}') \geq \mu(\mathcal{G}) - \delta$. Now at least one interval \mathcal{I}_k must satisfy

$$\frac{\mu(\mathcal{I}_k \cap \mathcal{G}^c)}{\mu(\mathcal{I}_k)} \leq \epsilon. \quad (69)$$

Otherwise, we would have

$$\begin{aligned} \mu(\mathcal{G}' \cap \mathcal{G}^c) &= \sum_{k=1}^K \mu(\mathcal{I}_k \cap \mathcal{G}^c) \\ &> \sum_{k=1}^K \epsilon\mu(\mathcal{I}_k) = \epsilon\mu(\mathcal{G}') \geq \epsilon(\mu(\mathcal{G}) - \delta) = \delta \end{aligned}$$

violating our assumption that $\mu(\mathcal{G}' \cap \mathcal{G}^c) \leq \delta$. So, let \mathcal{I}_k denote an interval that satisfies (69). Define $\gamma = 1/(2\mu(\mathcal{I}_k))$. Since

$D^+(\Lambda_{p_o}) > 0$, by Proposition 1 and (2), there exists $\tau \in \mathbb{R}$ such that

$$\#(\Lambda_{p_o} \cap B_\gamma(\tau)) \geq 2\gamma(D^+(\Lambda_{p_o})/2). \quad (70)$$

Define

$$X_{r_o}(f) = \begin{cases} e^{-j2\pi f\tau}/G_{p_o, r_o}(f), & \text{if } f \in \mathcal{I}_k \cap \mathcal{G} \\ 0, & \text{otherwise} \end{cases}$$

and $X_r(f) = 0$ for all $r \neq r_o$. Then, we clearly have

$$\|\mathbf{x}\|^2 \leq \epsilon\mu(\mathcal{I}_k). \quad (71)$$

Using (10), we conclude that

$$\begin{aligned} Y_{p_o}(f) &= e^{-j2\pi f\tau} \chi(f \in \mathcal{I}_k \cap \mathcal{G}) \\ &= e^{-j2\pi f\tau} [\chi(f \in \mathcal{I}_k) - \chi(f \in \mathcal{I}_k \cap \mathcal{G}^c)] \end{aligned}$$

whose inverse Fourier transform is

$$\begin{aligned} y_{p_o}(t) &= \mu(\mathcal{I}_k) \text{sinc}(\mu(\mathcal{I}_k)(t - \tau)) e^{-j2\pi f_0(t - \tau)} \\ &\quad - \phi_{\mathcal{I}_k \cap \mathcal{G}^c}(t - \tau), \quad t \in \mathbb{R} \end{aligned} \quad (72)$$

where $\text{sinc}(t) = \sin(\pi t)/(\pi t)$ and f_0 is the midpoint of \mathcal{I}_k . Note that

$$\sup_t |\phi_{\mathcal{I}_k \cap \mathcal{G}^c}(t)| \leq \mu(\mathcal{I}_k \cap \mathcal{G}^c) \leq \epsilon\mu(\mathcal{I}_k). \quad (73)$$

Then, for all t such that $|t - \tau| \leq \gamma = 1/(2\mu(\mathcal{I}_k))$, (72) and (73) imply that

$$|y_{p_o}(t)| \geq \mu(\mathcal{I}_k)(\text{sinc}(1/2) - \epsilon). \quad (74)$$

Using (70) and (74), we obtain

$$\begin{aligned} \sum_{p=1}^P \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 &\geq \sum_{n \in \mathbb{Z}} |y_{p_o}(\lambda_{np_o})|^2 \\ &\geq \sum_{\lambda \in \Lambda_{p_o} \cap B_\gamma(\tau)} |y_{p_o}(\lambda)|^2 \\ &\geq \frac{D^+(\Lambda_{p_o})}{2} 2\gamma [\mu(\mathcal{I}_k)(\text{sinc}(1/2) - \epsilon)]^2 \\ &= \frac{1}{2} D^+(\Lambda_{p_o}) \mu(\mathcal{I}_k) [\text{sinc}(1/2) - \epsilon]^2. \end{aligned}$$

Combining this result with (71), we obtain

$$\sum_{p=1}^P \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 \geq \frac{1}{2\epsilon} D^+(\Lambda_{p_o}) [\text{sinc}(1/2) - \epsilon]^2 \|\mathbf{x}\|^2.$$

Since $\epsilon > 0$ is arbitrary and $\mathbf{x} \in \mathcal{H}$ is nonzero, the above observation violates the second inequality of the stability condition (11), proving the necessity of (22). \square

D. Proof of Theorem 3

First note that the consistency condition implies that

$$\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : p \in \mathcal{P}, n \in \mathbb{Z}\} \quad (75)$$

is a Riesz–Fischer sequence in \mathcal{H} , where $\boldsymbol{\psi}_p$ is defined in (36). Let $\Pi \subseteq \mathcal{P}$ be a fixed subset. Consider the following two cases. First suppose that $\Pi \neq \emptyset$, $D^+(\Lambda_p) < \infty$ for all $p \in \Pi \neq \emptyset$, and that (24) holds. Then, we can define

$$K = \max_{p \in \Pi} C'(\Lambda_p) C(1) < \infty \quad (76)$$

where C' and C are quantities defined in Lemma 1. Since (24) holds, we can find a set \mathcal{D}_0 such that $\mu(\mathcal{D}_0) > 0$, and

$$\sigma_{\min}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) \leq \epsilon_0, \quad \forall f \in \mathcal{D}_0 \quad (77)$$

where $\epsilon_0 > 0$ is such that $K\epsilon_0^2 \leq a/4$ and a is the bound for the Riesz–Fischer sequence in (75). Assume without loss of

generality that $\mathcal{D}_0 \subseteq [\nu, \nu + 1]$ for some $\nu \in \mathbb{R}$. In the second case (either $\Pi = \emptyset$ or $D^+(\Lambda_p) = \infty$ for some $p \in \Pi$, or (24) does not hold), take $\mathcal{D}_0 = \emptyset$. Thus, (77) is satisfied in both cases. Let the rank of $\mathbf{G}_{\Pi, \mathcal{C}_f}^H(f)$ be denoted by

$$\rho(f) = \text{rank}(\mathbf{G}_{\Pi, \mathcal{C}_f}(f)) \quad (78)$$

and let the columns of $\mathbf{U}'(f) \in \mathbb{C}^{|\mathcal{C}_f| \times \rho(f)}$ form an orthonormal basis for the range space of $\mathbf{G}_{\Pi, \mathcal{C}_f}^H(f)$. Note that $\rho(f)$ is defined differently from $\rho(f)$ in the proof of Theorem 1. For $f \in \mathcal{D}_0$, let $\mathbf{U}''(f) \in \mathbb{C}^{|\mathcal{C}_f| \times 1}$ be a unit-norm right singular vector of $\mathbf{G}_{\Pi, \mathcal{C}_f}(f)$ corresponding to its smallest nonzero singular value. There is no loss of generality in assuming that the first column of $\mathbf{U}'(f)$ equals $\mathbf{U}''(f)$ for all $f \in \mathcal{D}_0$. Hence, for $f \in \mathcal{D}_0$, we can write

$$\mathbf{U}'(f) = [\mathbf{U}''(f) \ \mathbf{U}(f)]$$

for some $\mathbf{U}(f)$. For $f \notin \mathcal{D}_0$, let $\mathbf{U}(f) = \mathbf{U}'(f)$. The columns of $\mathbf{U}(f)$ are clearly orthonormal. Note that $\mathbf{U}'(f)$ and $\mathbf{U}''(f)$ can be assumed to be measurable functions. The matrix $\mathbf{U}(f)$ has r columns for $f \in \mathcal{G}_r$, where

$$\mathcal{G}_r = \{f : \rho(f) - \chi(f \in \mathcal{D}_0) = r\}, \quad r \in \mathcal{R}. \quad (79)$$

Each \mathcal{G}_r has finite measure. Since the sets $\{\mathcal{G}_r\}$ are disjoint, we can find, as in the proof of Theorem 1, a collection of disjoint intervals $\{\mathcal{I}_{rk} : r \in \mathcal{R}, k = 1, \dots, K_r\}$ and sets \mathcal{G}'_r as in (45) such that (46) and (47) hold. In the rest of the proof, assume that $q = q(r, k, m)$ for some invertible index-mapping as in the proof of Theorem 1.

Define $\mathbf{W}^r(f) \in \mathbb{C}^{R \times r}$, $\mathbf{S}_q(f)$, and Σ_q exactly as in (48)–(50), respectively. Also let \mathcal{H}_S be the closed subspace of \mathcal{H}_∞ defined as in (51). Using arguments similar to those in Theorem 1, we see that

$$\{\Theta_{\sigma_n q} \mathbf{s}_q : q \in \mathcal{Q}, n \in \mathbb{Z}\}$$

is an orthonormal Riesz basis for \mathcal{H}_S . In particular, it is a frame for \mathcal{H}_S . It is also easy to verify that \mathcal{H}_S is a shift-invariant subspace of \mathcal{H}_∞ .

Now, $\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : p \in \Pi, n \in \mathbb{Z}\}$, being a subcollection of the set in (75), is also a Riesz–Fischer sequence in \mathcal{H} . Let $\{c_{np}\}$, $p \in \Pi$ be some finite sequences. Then, (37) implies that

$$\max_{\mathbf{x} \in \mathcal{B}_{\mathcal{H}}} \left| \sum_{n,p} c_{np} y_p(\lambda_{np}) \right|^2 \geq a \sum_{n,p} |c_{np}|^2. \quad (80)$$

Let $\mathbf{x}^\circ \in \mathcal{B}_{\mathcal{H}}$ be the maximizer of the left-hand side of (80), and $\mathbf{Y}^\circ(f) = \mathbf{G}(f) \mathbf{X}^\circ(f)$, its corresponding MIMO channel output. Then

$$\left| \sum_{n,p} c_{np} y_p^\circ(\lambda_{np}) \right| \geq \sqrt{a} \|c\|. \quad (81)$$

Next, the subspace

$$\mathcal{H}_L = \{\mathbf{x} \in \mathcal{H} : \mathbf{X}_{\mathcal{C}_f}(f) = \mathbf{U}(f) \mathbf{U}^H(f) \mathbf{X}_{\mathcal{C}_f}(f) \text{ a.e.}\} \quad (82)$$

is closed and shift-invariant by the same argument as in the proof of Theorem 1. Note that

$$\mathbf{X}(f) = \mathbf{I}_{\bullet, \mathcal{C}_f} \mathbf{U}(f) \mathbf{U}^H(f) \mathbf{I}_{\mathcal{C}_f, \bullet} \mathbf{X}(f) \quad (83)$$

is an equivalent way of stating that $\mathbf{x} \in \mathcal{H}_L$ because $\mathbf{X}_{\mathcal{C}_f}(f)$ contains all the nonzero elements of $\mathbf{X}(f)$ for every $\mathbf{x} \in \mathcal{H}$. Define $\mathbf{x}' \in \mathcal{H}$ as follows:

$$\mathbf{X}'_{\mathcal{C}_f}(f) = \mathbf{U}(f) \mathbf{U}^H(f) \mathbf{X}^\circ_{\mathcal{C}_f}(f).$$

Evidently $\mathbf{x}' \in \mathcal{H}_L$ and $\|\mathbf{x}'\| \leq 1$ because $\mathbf{U}(f)$ has orthonormal columns. Let $\mathbf{Y}'(f) = \mathbf{G}(f) \mathbf{X}'(f)$ and $\mathbf{Y}''(f) = \mathbf{Y}^\circ(f) - \mathbf{Y}'(f)$, and recall that, for all $f \notin \mathcal{D}_0$, we have $\mathbf{U}(f) = \mathbf{U}'(f)$. Hence, using the definition of $\mathbf{U}'(f)$, we conclude that

$$\begin{aligned} \mathbf{Y}''_{\Pi}(f) &= \mathbf{G}_{\Pi, \mathcal{C}_f}(f) [\mathbf{X}^\circ_{\mathcal{C}_f}(f) - \mathbf{U}'(f) \mathbf{U}^H(f) \mathbf{X}^\circ_{\mathcal{C}_f}(f)] \\ &= \mathbf{0}, \quad \forall f \in \mathcal{D}_0. \end{aligned} \quad (84)$$

For $f \in \mathcal{D}_0$, we have

$$\mathbf{U}(f) \mathbf{U}^H(f) = \mathbf{U}'(f) \mathbf{U}^H(f) - \mathbf{U}''(f) \mathbf{U}''^H(f).$$

Hence,

$$\begin{aligned} \|\mathbf{Y}''_{\Pi}(f)\| &= \|\mathbf{G}_{\Pi, \mathcal{C}_f}(f) [\mathbf{X}^\circ_{\mathcal{C}_f}(f) - \mathbf{U}(f) \mathbf{U}^H(f) \mathbf{X}^\circ_{\mathcal{C}_f}(f)]\| \\ &= \|\mathbf{G}_{\Pi, \mathcal{C}_f}(f) \mathbf{U}''(f) \mathbf{U}''^H(f) \mathbf{X}^\circ_{\mathcal{C}_f}(f)\| \\ &\leq \epsilon_0 \|\mathbf{X}^\circ_{\mathcal{C}_f}(f)\|, \quad \forall f \in \mathcal{D}_0. \end{aligned} \quad (85)$$

Combining (84) and (85), and using $\|\mathbf{x}^\circ\| \leq 1$, we obtain

$$\int_{\mathbb{R}} \|\mathbf{Y}''_{\Pi}(f)\|^2 df \leq \begin{cases} \epsilon_0^2, & \text{if } \mu(\mathcal{D}_0) > 0 \\ 0, & \text{if } \mu(\mathcal{D}_0) = 0. \end{cases} \quad (86)$$

Recall that $\mathbf{Y}''_p(f)$ is supported on $\mathcal{D}_0 \subseteq [\nu, \nu + 1]$ for all $p \in \Pi$. Also, $\mu(\mathcal{D}_0) > 0$ implies that $D^+(\Lambda_p) < \infty$. Thus, we invoke Lemma 1 to get

$$\sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} |y_p''(\lambda_{np})|^2 \leq K \epsilon_0^2 \quad (87)$$

where K is the constant defined in (76). However, if $\mu(\mathcal{D}_0) = 0$, then $\mathbf{Y}''_{\Pi}(f) = \mathbf{0}$ from (86), and hence (87) holds trivially. In other words, (87) always holds. Using this and the Cauchy–Schwarz inequality, we conclude that

$$\left| \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} c_{np} y_p''(\lambda_{np}) \right| \leq \sqrt{K} \epsilon_0 \|c\|. \quad (88)$$

Recall that ϵ_0 is chosen so that $K \epsilon_0^2 \leq a/4$. Thus, combining (81) and (88) and noting that $y_p' = y_p^\circ - y_p''$, we obtain

$$\left| \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} c_{np} y_p'(\lambda_{np}) \right| \geq (\sqrt{a} - \sqrt{K} \epsilon_0) \|c\| \geq \frac{\sqrt{a}}{2} \|c\|.$$

Since the quantities $\{y_p'\}$ are the channel outputs corresponding to an input $\mathbf{x}' \in \mathcal{H}_L$ satisfying $\|\mathbf{x}'\| \leq 1$, we have

$$\max_{\mathbf{x}' \in \mathcal{B}_{\mathcal{H}_L}} \left| \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} c_{np} y_p'(\lambda_{np}) \right|^2 \geq \frac{a}{4} \|c\|^2. \quad (89)$$

Define $\mathbf{l}_p = P_{\mathcal{H}_L} \boldsymbol{\psi}_p$. Since \mathcal{H}_L is shift-invariant, we obtain $\Theta_{\lambda_{np}} \mathbf{l}_p \in \mathcal{H}_L$ and $y_p(\lambda_{np}) = \langle \mathbf{x}', \Theta_{\lambda_{np}} \mathbf{l}_p \rangle$ using the same argument as in (60) and (61). Therefore, (89) implies that

$$\left\| \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} c_{np} \Theta_{\lambda_{np}} \mathbf{l}_p \right\|^2 \geq \frac{a}{4} \|c\|^2$$

for all finite sequences $\{c_{np}\}$, $p \in \Pi$. Using (33), we conclude that $\{\Theta_{\lambda_{np}} \mathbf{l}_p : p \in \Pi, n \in \mathbb{Z}\}$ is a Riesz–Fischer sequence in \mathcal{H}_L with bound $a/4$.

To avoid confusion, we point out that the quantities associated with the frame are \mathcal{H}_S , Σ_q , \mathbf{s}_q , etc., while those associated with the Riesz–Fischer sequence are \mathcal{H}_L , Λ_p , \mathbf{l}_p , etc. This is opposite from the convention adopted in the proof of Theorem 1. In order to estimate the coefficients

$$\alpha_p = \sqrt{\frac{4}{a}} \sup_{\lambda \in \Lambda_p} \|\Theta_{\lambda} \mathbf{l}_p - P_{\mathcal{H}_S} \Theta_{\lambda} \mathbf{l}_p\|$$

of Theorem 4, we define $\mathbf{v}_p \in \mathcal{H}_\infty$ for $p \in \Pi$ as follows:

$$\mathbf{V}_p(f) = \begin{cases} \mathbf{L}_p(f), & \text{if } f \in \mathcal{G}'_r \cap \mathcal{G}_r, \text{ for some } r \\ \mathbf{0}, & \text{otherwise.} \end{cases} \quad (90)$$

Since $\mathbf{l}_p \in \mathcal{H}_L$, (83) implies that

$$\mathbf{L}_p(f) = \mathbf{I}_{\bullet, c_f} \mathbf{U}(f) \mathbf{U}^H(f) \mathbf{I}_{c_f, \bullet} \mathbf{L}_p(f).$$

Therefore, whenever $f \in \mathcal{G}'_r \cap \mathcal{G}_r$, we have

$$\begin{aligned} \mathbf{V}_p(f) &= \mathbf{I}_{\bullet, c_f} \mathbf{U}(f) \mathbf{U}^H(f) \mathbf{I}_{c_f, \bullet} \mathbf{L}_p(f) \\ &= \sum_{m=1}^r \mathbf{I}_{\bullet, c_f} \mathbf{U}_{\bullet, m}(f) [\mathbf{U}_{\bullet, m}^H(f) \mathbf{I}_{c_f, \bullet} \mathbf{L}_p(f)]. \end{aligned}$$

Combining this result with (45), (49), we can express $\mathbf{V}_p(f)$ as

$$\begin{aligned} \mathbf{V}_p(f) &= \sum_{r=1}^R \sum_{k=1}^{K_r} \sum_{m=1}^r \sqrt{\mu(\mathcal{I}_{rk})} \mathbf{S}_q(f) \\ &\quad \times [\mathbf{U}_{\bullet, m}^H(f) \mathbf{I}_{c_f, \bullet} \mathbf{L}_p(f)] \chi(f \in \mathcal{G}_r), \quad f \in \mathbb{R}. \end{aligned}$$

Since $[\mathbf{U}_{\bullet, m}^H(f) \mathbf{I}_{c_f, \bullet} \mathbf{L}_p(f)]$ is square-integrable and \mathcal{H}_S is shift-invariant, the time-domain expression for \mathbf{v}_p is obtained by first convolving each \mathbf{s}_q with a square-integrable function and then adding them together. Thus, $\mathbf{v}_p \in \mathcal{H}_S$, and using Proposition 2 we obtain

$$\begin{aligned} \alpha_p &= \sqrt{\frac{4}{a}} \sup_{\lambda \in \Lambda_p} \|\Theta_\lambda \mathbf{l}_p - P_{\mathcal{H}_S} \Theta_\lambda \mathbf{l}_p\| \\ &= \frac{2}{\sqrt{a}} \|\mathbf{l}_p - P_{\mathcal{H}_S} \mathbf{l}_p\| \leq \frac{2}{\sqrt{a}} \|\mathbf{l}_p - \mathbf{v}_p\|. \end{aligned}$$

Using Parseval's theorem and (90), we obtain the following estimate for α_p :

$$\begin{aligned} \alpha_p &\leq \frac{2}{\sqrt{a}} \left(\int_{\mathbb{R}} \|\mathbf{L}_p(f) - \mathbf{V}_p(f)\|^2 df \right)^{\frac{1}{2}} \\ &= \frac{2}{\sqrt{a}} \left(\sum_{r=1}^R \int_{\mathcal{G}'_r \cap \mathcal{G}_r} \|\mathbf{L}_p(f)\|^2 df \right)^{\frac{1}{2}}. \end{aligned}$$

This quantity can be made arbitrarily small for sufficiently small δ , because each $\mathbf{L}_p(f)$ is square-integrable, and $\mu(\mathcal{G}'_r \cap \mathcal{G}_r) \leq \delta/R^2$. Hence, for any $\epsilon > 0$ and sufficiently small $\delta > 0$, we can guarantee that $\alpha_p \leq \epsilon$. Applying Theorem 4, we obtain $D^+(\Lambda_p) < \infty$ for $p \in \Pi$ and

$$D^+(\Sigma_1, \dots, \Sigma_Q) \geq D^+(\{\Lambda_p : p \in \Pi\}) - \sum_{p \in \Pi} \alpha_p D^+(\Lambda_p). \quad (91)$$

Using the estimate for α_p and the fact that Σ_q has a uniform density of $\mu(\mathcal{I}_{rk})$ in (91), we obtain

$$\begin{aligned} D^+(\{\Lambda_p : p \in \Pi\}) &\leq \sum_{r,n,k} \mu(\mathcal{I}_{rk}) + \epsilon \sum_{p \in \Pi} D^+(\Lambda_p) \\ &= \sum_{r=1}^R r \mu(\mathcal{G}'_r) + \epsilon \sum_{p \in \Pi} D^+(\Lambda_p) \\ &\leq \sum_{r=1}^R r \mu(\mathcal{G}_r) + \delta + \epsilon \sum_{p \in \Pi} D^+(\Lambda_p). \quad (92) \end{aligned}$$

Next, (78), (79), and the definition of the Lebesgue integral yield

$$\begin{aligned} \sum_{r=1}^R r \mu(\mathcal{G}_r) &= \int_{\mathbb{R}} [\rho(f) - \chi(f \in \mathcal{D}_0)] df \\ &= \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi, c_f}(f)) df - \mu(\mathcal{D}_0). \quad (93) \end{aligned}$$

Combining (92) and (93), and letting $\delta, \epsilon \rightarrow 0$, we obtain

$$D^+(\{\Lambda_p : p \in \Pi\}) \leq \int_{\mathbb{R}} \text{rank}(\mathbf{G}_{\Pi, c_f}(f)) df - \mu(\mathcal{D}_0).$$

This proves (23). We have already demonstrated that $D^+(\Lambda_p) < \infty$. Therefore, if (24) holds and $\Pi \neq \emptyset$, we have $\mu(\mathcal{D}_0) > 0$, implying that the inequality in (23) is strict. \square

V. SUMMARY

We formulated the MIMO sampling scheme, and defined stable sampling and consistent reconstruction. These are generalizations of stable sampling and interpolation for classical sampling. We also introduced notions of upper and lower sampling densities applicable to *collections* of sampling sets. We derived necessary density conditions for stable sampling and consistent reconstruction in the MIMO setting. For stable sampling, we find that a family of $2^P - 1$ bounds hold—a lower bound on the joint lower density of each nonempty set of P output sampling sets. Similarly, we find that a family of $2^P - 1$ bounds hold for the consistency problem which are upper bounds on the joint upper densities of the sampling sets. These bounds generalize Landau's necessary density results for classical sampling. Since the MIMO sampling scheme is extremely general, and encompasses various sampling schemes such as Papoulis' generalized sampling, and multicore or periodic nonuniform sampling as special cases, we automatically have necessary conditions for all these sampling schemes as well.

REFERENCES

- [1] B. R. Petersen and D. D. Falconer, "Suppression of adjacent-channel, co-channel, and intersymbol interference by equalizers and linear combiners," *IEEE Trans. Commun.*, vol. 42, pp. 3109–3118, Dec. 1994.
- [2] J. Yang and S. Roy, "On joint receiver and transmitted optimization for multiple-input–multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, pp. 3221–3231, Dec. 1994.
- [3] G. R. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [4] L. Ye and K. J. R. Liu, "Adaptive blind source separation and equalization for multiple-input/multiple-output systems," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2864–2876, Nov. 1998.
- [5] W. Zishun and J. D. Z. Chen, "Blind separation of slow waves and spikes from gastrointestinal myoelectrical recordings," *IEEE Trans. Inform. Technol. Biomed.*, vol. 5, no. 2, pp. 133–137, June 2001.
- [6] V. Zarzoso and A. K. Nandi, "Noninvasive fetal electrocardiogram extraction: Blind separation versus adaptive noise cancellation," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 12–18, Jan. 2001.
- [7] P. a. Voois and J. M. Cioffi, "Multichannel signal processing for multiple-head digital magnetic recording," *IEEE Trans. Magn.*, vol. 30, no. 6, pp. 5100–5114, Nov. 1994.
- [8] K.-C. Yen and Y. Zhao, "Adaptive co-channel speech separation and recognition," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 138–151, Mar. 1999.
- [9] A. González and J. J. Lopéz, "Fast transversal filters for deconvolution in multichannel sound reproduction," *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 429–440, May 2001.

- [10] J. Idier and Y. Goussard, "Multichannel seismic deconvolution," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 961–979, Sept. 1993.
- [11] G. Harikumar and Y. Bresler, "Exact image deconvolution from multiple FIR blurs," *IEEE Trans. Image Processing*, vol. 8, pp. 846–862, June 1999.
- [12] G. B. Giannakis and R. W. Heath Jr., "Blind identification of multichannel FIR blurs and perfect image restoration," *IEEE Trans. Image Processing*, vol. 9, pp. 1877–1896, Nov. 2000.
- [13] R. J. Papoulis, "Generalized sampling expansions," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 652–654, Nov. 1977.
- [14] D. Seidner and M. Feder, "Vector sampling expansions," *IEEE Trans. Signal Processing*, vol. 48, pp. 1401–1416, May 2000.
- [15] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, 1967.
- [16] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proc. IEEE*, vol. 55, pp. 1701–1706, Oct. 1967.
- [17] K. Gröchenig and H. Razafinjato, "On Landau's necessary density conditions for sampling and interpolation of bandlimited functions," *J. London Math. Soc.*, vol. 2, no. 54, pp. 557–565, 1996.
- [18] R. Venkataramani and Y. Bresler, "Sampling theorems for uniform and periodic nonuniform MIMO sampling of multiband signals," *IEEE Trans. Signal Processing*, vol. 51, pp. 3152–3163, Dec. 2003.
- [19] —, "Filter design for MIMO sampling and reconstruction," *IEEE Trans. Signal Processing*, vol. 51, pp. 3164–3176, Dec. 2003.
- [20] M. Unser and J. Zerubia, "Generalized sampling: Stability and analysis," *IEEE Trans. Signal Processing*, vol. 45, pp. 2941–2950, Dec. 1997.
- [21] J. Benedetto and P. J. S. G. Ferreira, Eds., *Modern Sampling Theory: Mathematics and Applications*. Boston, MA: Birkhäuser, 2001.
- [22] R. Venkataramani, "Sub-Nyquist multicoset and MIMO sampling: Perfect reconstruction, performance analysis, and necessary density conditions," Ph.D. dissertation, Univ. Illinois at Urbana-Champaign, Urbana, 2001.
- [23] R. M. Young, *An Introduction to Nonharmonic Fourier Analysis*. San Diego, CA: Academic, 2001.
- [24] J. R. Higgins, *Sampling Theory in Fourier and Signals Analysis Foundations*. New York: Oxford Science Pub., 1996.